84345

STIC-Biotech/ChemLib

From: Nickol, Gary
Sent: Wednesday, January 15, 2003 12:14 PM
To: STIC-Biotech/ChemLib
Subject: 09/455,486

Please search and interference search the amino acids of SEQ ID NO:6.

Thanks!

Gary Nickol
CM1, AU:1642
Room 8D17, Mailbox 8E12
(703) 305-7143

CRF

Point of Contact
P. Sheppard
Telephone number: (703) 308-4499

	TYPE OF SEARCH:	VENDOR/COST (where applic.)
Searcher:_____	NA Sequences:_____	STN:_____
Phone:_____	AA Sequences:_____	DIALOG:_____
Location:_____	Structures:_____	Questel/Orbit:_____
Date Picked Up:_____	Bibliographic:_____	DRLink:_____
Date Completed: 1/15/03	Litigation:_____	Lexis/Nexis:_____
Searcher Prep/Review:_____	Full text:_____	Sequence Sys.:_____
Clerical:_____	Patent Family:_____	WWW/Internet:_____
Online time:_____	Other:_____	Other (specify):_____

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: January 17, 2003, 18:58:14 ; Search time 41 Seconds
(without alignments)
1475.507 Million cell updates/sec

Title: US-09-455-486-6
Perfect score: 2351
Sequence: 1 MESISMMGSPKSLSETCLPN.....ALVLPSIVILDLLQLCRYPD 454

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2351	100.0	454	22	AAU10188	Human ORF2 of Six-
2	2351	100.0	454	22	AAE02781	Human six transmem
3	2351	100.0	454	23	ABG61933	Prostate cancer-as
4	2351	100.0	454	23	AAU80190	Human PUMPCn prote
5	2351	100.0	454	23	AAU76538	Tumour-associated
6	2290	97.4	490	22	AAU10187	Human Six-Transmem
7	2036	86.6	419	22	AAU10189	Human ORF3 of Six-
8	1972.5	83.9	576	22	ABG12306	Novel human diagno
9	1972.5	83.9	1273	22	ABG00113	Novel human diagno
10	1738	73.9	450	22	AAE02841	Human STEAP-2 prot
11	1316	56.0	488	22	AAB49483	Rat p-HYDE. Rattu
12	1316	56.0	526	23	ABB83365	Murine Tumour Supp
13	1272	54.1	488	22	AAB85775	Human drug metabol
14	1272	54.1	488	23	ABB83366	Human Tumour Suppr
15	1269	54.0	488	22	AAU10220	Human Six-Transmem
16	1269	54.0	488	22	AAU04564	Human G-protein co
17	1269	54.0	488	22	AAB93224	Human protein sequ
18	1245.5	53.0	487	22	AAB49481	Human p-HYDE. Hom
19	1088	46.3	459	22	AAB74715	Human membrane ass
20	1087	46.2	459	22	AAE02782	Human six transmem
21	1085	46.2	459	22	AAU10190	Human Six-Transmem
22	1082	46.0	458	22	AAE02636	Human STEAP-3 prot
23	1074.5	45.7	456	22	AAB49482	Human p-HYDE 40.
24	901	38.3	173	21	AAY58195	Human STRAP-2 prot
25	844	35.9	179	22	ABB31810	Peptide #4461 enco
26	844	35.9	179	22	ABB37041	Peptide #4547 enco
27	844	35.9	179	22	ABB22356	Protein #4355 enco
28	844	35.9	179	22	AAM57768	Human brain expres
29	844	35.9	179	22	AAM70183	Human bone marrow
30	844	35.9	179	22	AAM18006	Peptide #4440 enco
31	844	35.9	179	22	AAM30517	Peptide #4554 enco
32	844	35.9	179	22	AAM05646	Peptide #4328 enco
33	844	35.9	179	23	ABG39817	Human peptide enco
34	768	32.7	237	22	AAU04565	Human G-protein co
35	736	31.3	141	21	AAY52589	Human prostate gro
36	717	30.5	339	21	AAY58194	Human STRAP-1 prot
37	717	30.5	339	22	AAU69927	Human prostate cDN
38	717	30.5	339	22	AAM78845	Human protein SEQ
39	717	30.5	339	22	AAM01282	P789P amino acid s
40	717	30.5	339	23	ABG61813	Prostate cancer-as
41	717	30.5	339	23	ABB95387	Human P789P protei
42	717	30.5	375	22	AAE02780	Human six transmem
43	704	29.9	374	22	AAM79829	Human protein SEQ
44	695	29.6	339	20	AAW86309	Kidney injury asso
45	435	18.5	129	22	AAB75314	Gene 20 human secr

ALIGNMENTS

RESULT 1
AAU10188
ID AAU10188 standard; Protein; 454 AA.
XX
AC AAU10188;
XX
DT 16-JAN-2002 (first entry)
XX
DE Human ORF2 of Six-Transmembrane Protein of Prostate 1.
XX
KW Human; Six-Transmembrane Protein of Prostate 1; STMP1; prostate cancer;
KW benign prostatic hyperplasia; acute prostatitis; testicular cancer;
KW cryptorchidism; testicular disorder; proliferative disorder; lymphoma;
KW leukaemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer;
KW liver cancer; lung cancer; cytostatic; ORF2.
XX
OS Homo sapiens.
XX
PN WO200172962-A2.
XX
PD 04-OCT-2001.
XX
PF 23-MAR-2001; 2001WO-US09410.
XX
PR 24-MAR-2000; 2000US-191929P.
XX
PA (SAAT/) SAATCIOGLU F.
XX
PI Saatcioglu F;
XX
DR WPI; 2001-662926/76.
DR N-PSDB; AAS15810, AAS15811.
XX

```
PT   New polynucleotide for the diagnosis, prevention and treatment for
PT   prostate and testis disorders, particularly prostate cancer, comprises
PT   prostate-specific or testis-specific nucleic acids  -
XX
PS   Claim 1; Fig 4H; 114pp; English.
XX
CC   The invention relates to substantially pure prostate-specific or
CC   testis-specific polypeptides and the nucleic acids encoding them.
CC   Also included are vectors and host cells expressing the proteins, a
CC   transgenic animal expressing the protein, antibodies against the
CC   proteins, probes for detecting the nucleic acids, antisense molecules
CC   for the nucleic acids and methods of isolating modulators of the
CC   proteins. Compounds that modulate the prostate specific or testis
CC   specific polypeptide are useful to diagnose, prevent or treat disorders
CC   of the testis or prostate particularly prostate cancer, benign
CC   prostatic hyperplasia, acute prostatitis, testicular cancer,
CC   cryptorchidism, undescended, retractile, ascending or vanished
CC   testis. Other proliferative disorders for which the modulators may be
CC   used include lymphoma, leukaemia, melanoma, ovarian cancer, breast
CC   cancer, pancreatic cancer, liver cancer and lung cancer. The
CC   present sequence is prostate specific protein, Six-Transmembrane
CC   Protein of Prostate 1, STMP1, ORF2.
XX
SQ   Sequence   454 AA;

  Query Match             100.0%;  Score 2351;  DB 22;  Length 454;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-244;
  Matches  454;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MESISMMGSPKSLSETCLPNGINGIKDARKVTVGVIGSGDFAKSLTIRLIRCGYHVVIGS 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MESISMMGSPKSLSETCLPNGINGIKDARKVTVGVIGSGDFAKSLTIRLIRCGYHVVIGS 60

Qy      61 RNPKFASEFFPHVVDVTHHEDALTKTNIIFVAIHREHYTSLWDLRHLLVGKILIDVSNNM 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 RNPKFASEFFPHVVDVTHHEDALTKTNIIFVAIHREHYTSLWDLRHLLVGKILIDVSNNM 120

Qy     121 RINQYPESNAEYLASLFPDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIE 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 RINQYPESNAEYLASLFPDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIE 180

Qy     181 LARQLNFIPIDLGSLSSAREIENLPLRLFTLWRGPVVVAISLATFFFLYSFVRDVIHPYA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 LARQLNFIPIDLGSLSSAREIENLPLRLFTLWRGPVVVAISLATFFFLYSFVRDVIHPYA 240

Qy     241 RNQQSDFYKIPIEIVNKTLPIVAITLLSLVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQ 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 RNQQSDFYKIPIEIVNKTLPIVAITLLSLVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQ 300

Qy     301 CRKQLGLLSFFFAMVHVAYSLCLPMRRSERYLFLNMAYQQVHANIENSWNEEEVWRIEMY 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CRKQLGLLSFFFAMVHVAYSLCLPMRRSERYLFLNMAYQQVHANIENSWNEEEVWRIEMY 360

Qy     361 ISFGIMSLGLLSLLAVTSIPSVSNALNWREFSFIQSTLGYVALLISTFHVLIYGWKRAFE 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 ISFGIMSLGLLSLLAVTSIPSVSNALNWREFSFIQSTLGYVALLISTFHVLIYGWKRAFE 420

Qy     421 EEYYRFYTPPNFVLALVLPSIVILDLLQLCRYPD 454
          |||||||||||||||||||||||||||||||||||
Db     421 EEYYRFYTPPNFVLALVLPSIVILDLLQLCRYPD 454


RESULT 2
AAE02781
ID   AAE02781 standard; Protein; 454 AA.
XX
AC   AAE02781;
XX
DT   06-AUG-2001  (first entry)
XX
DE   Human six transmembrane epithelial antigen of prostate (STEAP)-2 protein.
XX
KW   Human; cytostatic; antiproliferative; vaccine; gene therapy;
KW   six transmembrane epithelial antigen of the prostate-2; STEAP-2;
KW   chromosome 7q21; cancer; prostate; colon; bladder; lung; ovarian;
KW   pancreatic.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          100..108
FT                   /label= HLA-A2_binding_peptide #5
FT   Region          153..165
FT                   /label= Immunogenic_peptide #1
FT   Domain          207..228
FT                   /label= Transmembrane_domain #1
FT   Region          227..235
FT                   /label= HLA-A2_binding_peptide #1
FT   Domain          255..277
FT                   /label= Transmembrane_domain #2
FT   Domain          304..325
FT                   /label= Transmembrane_domain #3
FT   Region          306..314
FT                   /label= HLA-A2_binding_peptide #4
FT   Region          307..315
FT                   /label= HLA-A2_binding_peptide #3
FT   Region          345..358
FT                   /label= Immunogenic_peptide #2
FT   Domain          359..381
FT                   /label= Transmembrane_domain #4
FT   Domain          393..415
FT                   /label= Transmembrane_domain #5
FT   Region          402..410
FT                   /label= HLA-A2_binding_peptide #2
FT   Domain          428..450
FT                   /label= Transmembrane_domain #6
XX
PN   WO200140276-A2.
XX
PD   07-JUN-2001.
XX
PF   06-DEC-2000; 2000WO-US33040.
XX
PR   06-DEC-1999;   99US-0455486.
XX
PA   (UROG-) UROGENESYS INC.
XX
PI   Afar DEH,  Hubert RS,  Raitano AB,  Saffran DC,  Mitchell SC,  Faris M;
PI   Jakobovits A;
XX
DR   WPI; 2001-367804/38.
DR   N-PSDB; AAD07072.
XX
PT   New STEAP (six transmembrane epithelial antigen of the prostate)
PT   proteins, expressed in human cancers, useful for detecting and treating
PT   cancer -
XX
PS   Claim 1; Fig 9A-9C; 187pp; English.
XX
CC   The present sequence is human six transmembrane epithelial antigen of
CC   the prostate (STEAP)-2 protein. STEAP is a member of cell
CC   surface serpentine transmembrane antigens. STEAP-2 gene is located on
CC   chromosome 7q21 and is used in gene therapy. Inhibiting the development
CC   or progression of a cancer (eg. prostate, colon, bladder, lung, ovarian
CC   and pancreatic) expressing STEAP or inhibiting growth or killing cells
CC   expressing STEAP in a patient, comprises administering a vaccine
CC   composition to the patient. Treating a patient with a cancer that
CC   expresses STEAP, or inhibiting growth or killing cells expressing STEAP,
CC   comprises administering to the patient a vector encoding single chain
CC   monoclonal antibody that comprises the variable domains of the heavy and
CC   light chains of the monoclonal antibody that specifically binds to STEAP,
CC   such that the vector delivers the single chain monoclonal antibody coding
CC   sequence to the cancer cells and the encoded single chain monoclonal
CC   antibody is expressed intracellularly.
CC   Note: This sequence is stated to be the same as that being shown as
CC   SEQ ID NO:8 (AAE02841) in sequence listing of the specification. However
```

```
CC   both the sequences differ at several positions.
XX
SQ   Sequence   454 AA;

  Query Match            100.0%;  Score 2351;  DB 22;  Length 454;
  Best Local Similarity  100.0%;  Pred. No. 1.3e-244;
  Matches  454;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MESISMMGSPKSLSETCLPNGINGIKDARKVTVGVIGSGDFAKSLTIRLIRCGYHVVIGS 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MESISMMGSPKSLSETCLPNGINGIKDARKVTVGVIGSGDFAKSLTIRLIRCGYHVVIGS 60

Qy     61 RNPKFASEFFPHVVDVTHHEDALTKTNIIFVAIHREHYTSLWDLRHLLVGKILIDVSNNM 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 RNPKFASEFFPHVVDVTHHEDALTKTNIIFVAIHREHYTSLWDLRHLLVGKILIDVSNNM 120

Qy    121 RINQYPESNAEYLASLFPDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIE 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 RINQYPESNAEYLASLFPDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIE 180

Qy    181 LARQLNFIPIDLGSLSSAREIENLPLRLFTLWRGPVVVAISLATFFFLYSFVRDVIHPYA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 LARQLNFIPIDLGSLSSAREIENLPLRLFTLWRGPVVVAISLATFFFLYSFVRDVIHPYA 240

Qy    241 RNQQSDFYKIPIEIVNKTLPIVAITLLSLVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQ 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 RNQQSDFYKIPIEIVNKTLPIVAITLLSLVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQ 300

Qy    301 CRKQLGLLSFFFAMVHVAYSLCLPMRRSERYLFLNMAYQQVHANIENSWNEEEVWRIEMY 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CRKQLGLLSFFFAMVHVAYSLCLPMRRSERYLFLNMAYQQVHANIENSWNEEEVWRIEMY 360

Qy    361 ISFGIMSLGLLSLLAVTSIPSVSNALNWREFSFIQSTLGYVALLISTFHVLIYGWKRAFE 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ISFGIMSLGLLSLLAVTSIPSVSNALNWREFSFIQSTLGYVALLISTFHVLIYGWKRAFE 420

Qy    421 EEYYRFYTPPNFVLALVLPSIVILDLLQLCRYPD 454
          |||||||||||||||||||||||||||||||||||
Db    421 EEYYRFYTPPNFVLALVLPSIVILDLLQLCRYPD 454


RESULT 3
ABG61933
ID   ABG61933 standard; Protein; 454 AA.
XX
AC   ABG61933;
XX
DT   15-AUG-2002  (first entry)
XX
DE   Prostate cancer-associated protein #134.
XX
KW   Prostate cancer; prostate tumour tissue; human; mammal; cytostatic.
XX
OS   Mammalia.
XX
PN   WO200230268-A2.
XX
PD   18-APR-2002.
XX
PF   12-OCT-2001; 2001WO-US32045.
XX
PR   13-OCT-2000; 2000US-0687576.
PR   08-DEC-2000; 2000US-0733288.
PR   08-DEC-2000; 2000US-0733742.
PR   24-JAN-2001; 2001US-263957P.
PR   16-MAR-2001; 2001US-276791P.
PR   16-MAR-2001; 2001US-276888P.
PR   06-APR-2001; 2001US-281922P.
PR   24-APR-2001; 2001US-286214P.
PR   30-APR-2001; 2001US-0847046.
PR   04-MAY-2001; 2001US-288589P.
XX
PA   (EOSB-) EOS BIOTECHNOLOGY INC.
XX
PI   Gish KC,  Mack DH,  Wilson KE,  Afar D,  Hevezi P;
XX
DR   WPI; 2002-471335/50.
DR   N-PSDB; ABK92252.
XX
PT   Detecting a prostate cancer-associated transcript in a cell in a
PT   patient, useful for diagnosing prostate cancer (PC) or screening
PT   modulators of PC, by determining if prostate cancer-associated genes
PT   are expressed in a prostate tissue -
XX
PS   Claim 27; Page 416; 436pp; English.
XX
CC   The present invention relates to methods of detecting a prostate
CC   cancer-associated transcript in a cell from a patient. The method
CC   comprises contacting a biological sample from the patient with
CC   prostate cancer-associated polynucleotides (designated PC genes) that
CC   selectively hybridise to a sequence that is at least 80% identical
CC   to them. The prostate cancer-associated polynucleotide sequences
CC   are differentially expressed in prostate tumour tissue or in
CC   prostate cancer and are derived from the tissues of various
CC   organisms such as humans or other mammals (e.g. mice, sheep and dogs).
CC   The methods of the invention are useful for diagnosing and treating
CC   prostate cancer in mammals. The prostate cancer-associated genes are
CC   useful for diagnosing or treating prostate cancer, as well as for
CC   identifying modulators of prostate cancer or agents that inhibit
CC   prostate cancer. The nucleic acid sequences are particularly useful
CC   in gene therapy, as a vaccine or in antisense applications.
CC   ABG61800-ABG61944 represent prostate cancer-associated proteins.
XX
SQ   Sequence   454 AA;

  Query Match            100.0%;  Score 2351;  DB 23;  Length 454;
  Best Local Similarity  100.0%;  Pred. No. 1.3e-244;
  Matches  454;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MESISMMGSPKSLSETCLPNGINGIKDARKVTVGVIGSGDFAKSLTIRLIRCGYHVVIGS 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MESISMMGSPKSLSETCLPNGINGIKDARKVTVGVIGSGDFAKSLTIRLIRCGYHVVIGS 60

Qy     61 RNPKFASEFFPHVVDVTHHEDALTKTNIIFVAIHREHYTSLWDLRHLLVGKILIDVSNNM 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 RNPKFASEFFPHVVDVTHHEDALTKTNIIFVAIHREHYTSLWDLRHLLVGKILIDVSNNM 120

Qy    121 RINQYPESNAEYLASLFPDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIE 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 RINQYPESNAEYLASLFPDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIE 180

Qy    181 LARQLNFIPIDLGSLSSAREIENLPLRLFTLWRGPVVVAISLATFFFLYSFVRDVIHPYA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 LARQLNFIPIDLGSLSSAREIENLPLRLFTLWRGPVVVAISLATFFFLYSFVRDVIHPYA 240

Qy    241 RNQQSDFYKIPIEIVNKTLPIVAITLLSLVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQ 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 RNQQSDFYKIPIEIVNKTLPIVAITLLSLVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQ 300

Qy    301 CRKQLGLLSFFFAMVHVAYSLCLPMRRSERYLFLNMAYQQVHANIENSWNEEEVWRIEMY 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CRKQLGLLSFFFAMVHVAYSLCLPMRRSERYLFLNMAYQQVHANIENSWNEEEVWRIEMY 360

Qy    361 ISFGIMSLGLLSLLAVTSIPSVSNALNWREFSFIQSTLGYVALLISTFHVLIYGWKRAFE 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ISFGIMSLGLLSLLAVTSIPSVSNALNWREFSFIQSTLGYVALLISTFHVLIYGWKRAFE 420

Qy    421 EEYYRFYTPPNFVLALVLPSIVILDLLQLCRYPD 454
          |||||||||||||||||||||||||||||||||||
Db    421 EEYYRFYTPPNFVLALVLPSIVILDLLQLCRYPD 454


RESULT 4
AAU80190
ID   AAU80190 standard; Protein; 454 AA.
XX
```

```
AC   AAU80190;
XX
DT   15-JUL-2002  (first entry)
XX
DE   Human PUMPCn protein, PRO23203.
XX
KW   Human; protein upregulated in metastatic prostate cancer; immunogen;
KW   PUMPCn; PRO23203; prostate cancer; cytostatic; gene therapy; transgenic;
KW   androgen independent prostate cancer; DNA microarray.
XX
OS   Homo sapiens.
XX
PN   WO200226822-A2.
XX
PD   04-APR-2002.
XX
PF   26-SEP-2001; 2001WO-US30290.
XX
PR   26-SEP-2000; 2000US-235451P.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Devaux B,  Eberhard D,  Goddard A,  Godowski PJ,  Grimaldi JC;
PI   Hillan KJ,  Watanabe CK,  Wood WI,  Yansura DG,  Zhang Z;
XX
DR   WPI; 2002-383270/41.
DR   N-PSDB; ABK50391.
XX
PT   New polypeptide termed protein upregulated in metastatic prostate
PT   cancer and encoding polynucleotides, useful for identifying polypeptide
PT   antagonists for treating prostate cancer
XX
PS   Claim 23; Fig 2; 137pp; English.
XX
CC   The invention relates to an isolated human protein upregulated
CC   in metastatic prostate cancer (PUMPCn) PRO23203 polypeptide, a
CC   sequence 80% identical to PRO23203 and the sequence as encoded by cDNA
CC   insert of the vector deposited as ATCC Deposit No. PTA-2513 (DNA185171-
CC   2994) on 26/9/2000. Also included are the polynucleotide encoding
CC   the protein (or a DNA sequence 80% identical to the polynucleotide
CC   and one that hybridises to complement of the polynucleotide), a vector
CC   comprising the polynucleotide, a polynucleotide deposited with ATCC under
CC   accession number PTA-2513 (DNA185171-2994), a host cell comprising the
CC   vector, preparation of PRO23203, a chimeric molecule comprising PRO23203
CC   fused to a heterologous amino acid sequence, an anti-PRO23203 antibody,
CC   an agonist/antagonist of PRO23203, and diagnosing the presence of
CC   prostate cancer in a mammal by: (a) contacting a microarray diagnostic
CC   with a DNA185171-2994 probe, detecting and quantifying hybridisation of
CC   DNA185171-2994 probe in prostate cancer tissue compared with normal
CC   tissue and determining if DNA185171-2994 is overexpressed; or
CC   (b) contacting a tissue of the mammal with an anti-PRO23203 antibody
CC   and detecting the binding of the antibody to a component of the tissue,
CC   where binding is indicative of the presence of prostate cancer in the
CC   mammal. The antibody is useful for treating prostate cancer in mammal
CC   which is androgen independent prostate cancer, that has metastasised to
CC   another portion of the body, where the antibody is not conjugated with a
CC   cytotoxic agent and the method further comprises administering a
CC   chemotherapeutic agent to the mammal and for diagnosing the presence of
CC   prostate cancer in an mammal. The PRO23203 polynucleotide is useful for
CC   generating either transgenic animals or knock out animals which in turn
CC   are useful in the development and screening of therapeutically useful
CC   reagents or for use in gene therapy and for chromosome identification.
CC   PRO23203 and polynucleotide may also be used for tissue typing and
CC   PRO23203 may also be employed as a therapeutic agents and for screening
CC   compounds to identify those that mimic the PRO23203 polypeptides
CC   (agonists) or prevent the effect of the PRO23203 polypeptide
CC   (antagonists). PRO23203 is further useful for the affinity purification
CC   of PRO23203 from recombinant cell culture or natural sources. The
CC   present sequence represents PRO23203.
XX
SQ   Sequence   454 AA;

 Query Match             100.0%;  Score 2351;  DB 23;  Length 454;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-244;
  Matches  454;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MESISMMGSPKSLSETCLPNGINGIKDARKVTVGVIGSGDFAKSLTIRLIRCGYHVVIGS 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MESISMMGSPKSLSETCLPNGINGIKDARKVTVGVIGSGDFAKSLTIRLIRCGYHVVIGS 60

Qy      61 RNPKFASEFFPHVVDVTHHEDALTKTNIIFVAIHREHYTSLWDLRHLLVGKILIDVSNNM 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 RNPKFASEFFPHVVDVTHHEDALTKTNIIFVAIHREHYTSLWDLRHLLVGKILIDVSNNM 120

Qy     121 RINQYPESNAEYLASLFPDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIE 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 RINQYPESNAEYLASLFPDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIE 180

Qy     181 LARQLNFIPIDLGSLSSAREIENLPLRLFTLWRGPVVVAISLATFFFLYSFVRDVIHPYA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 LARQLNFIPIDLGSLSSAREIENLPLRLFTLWRGPVVVAISLATFFFLYSFVRDVIHPYA 240

Qy     241 RNQQSDFYKIPIEIVNKTLPIVAITLLSLVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQ 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 RNQQSDFYKIPIEIVNKTLPIVAITLLSLVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQ 300

Qy     301 CRKQLGLLSFFFAMVHVAYSLCLPMRRSERYLFLNMAYQQVHANIENSWNEEEVWRIEMY 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CRKQLGLLSFFFAMVHVAYSLCLPMRRSERYLFLNMAYQQVHANIENSWNEEEVWRIEMY 360

Qy     361 ISFGIMSLGLLSLLAVTSIPSVSNALNWREFSFIQSTLGYVALLISTFHVLIYGWKRAFE 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 ISFGIMSLGLLSLLAVTSIPSVSNALNWREFSFIQSTLGYVALLISTFHVLIYGWKRAFE 420

Qy     421 EEYYRFYTPPNFVLALVLPSIVILDLLQLCRYPD 454
           |||||||||||||||||||||||||||||||||
Db     421 EEYYRFYTPPNFVLALVLPSIVILDLLQLCRYPD 454


RESULT 5
AAU76538
ID   AAU76538 standard; Protein; 454 AA.
XX
AC   AAU76538;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Tumour-associated antigenic target protein, TAT138.
XX
KW   TAT138; Tumour-associated Antigenic Target; tumour;
KW   breast cancer; colorectal cancer; lung cancer; ovarian cancer;
KW   central nervous system cancer; liver cancer; bladder cancer; melanoma;
KW   pancreatic cancer; leukaemia; gene therapy.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   8..13
FT                   /note= "N-myristoylation site"
FT   Modified-site   24..29
FT                   /note= "N-myristoylation site"
FT   Region          29..32
FT                   /note= "cAMP- and cGMP-dependent protein kinase
FT                    phosphorylation site"
FT   Modified-site   34..39
FT                   /note= "N-myristoylation site"
FT   Modified-site   193..198
FT                   /note= "N-myristoylation site"
FT   Domain          210..230
FT                   /note= "Transmembrane domain"
FT   Domain          257..277
FT                   /note= "Transmembrane domain"
FT   Modified-site   256..259
FT                   /note= "Asn is N-glycosylated"
FT   Modified-site   274..279
FT                   /note= "N-myristoylation site"
```

```
FT   Domain          299..319
FT                   /note= "Transmembrane domain"
FT   Domain          259..379
FT                   /note= "Transmembrane domain"
FT   Domain          393..413
FT                   /note= "Transmembrane domain"
FT   Modified-site   416..424
FT                   /note= "Tyrosine kinase phosphorylation site"
FT   Domain          428..448
FT                   /note= "Transmembrane domain"
XX
PN   WO200216429-A2.
XX
PD   28-FEB-2002.
XX
PF   22-JUN-2001; 2001WO-US20118.
XX
PR   24-AUG-2000; 2000WO-US23328.
PR   26-SEP-2000; 2000US-235451P.
PR   01-DEC-2000; 2000WO-US32678.
PR   28-FEB-2001; 2001WO-US06520.
PR   01-MAR-2001; 2001WO-US06666.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Goddard A, Godowski PJ, Gurney AL, Hillan KJ, Polakis P, Smith V;
PI   Wood WI, Wu TD, Zhang Z;
XX
DR   WPI; 2002-280917/32.
DR   N-PSDB; ABK11093.
XX
PT   Novel isolated tumour-associated antigenic target polypeptides which are
PT   useful as targets for cancer therapy and diagnosis in mammals -
XX
PS   Claim 12; Fig 10; 121pp; English.
XX
CC   The invention relates to an isolated tumour-associated antigenic target
CC   polypeptide (TAT) (I), specifically TAT134-TAT138 polypeptides, and the
CC   polynucleotides (II) encoding them. (II) is useful for diagnosing the
CC   presence of a tumour in a mammal, where the level of expression of (II)
CC   is indicative on the presence of tumour in the mammal from which the
CC   test sample was obtained. Antibody to (I) is useful for killing a cancer
CC   cell (e.g. breast cancer cell, a colorectal cancer cell, a lung cancer
CC   cell, an ovarian cancer cell, a central nervous system (CNS) cancer cell,
CC   a liver cancer cell, a bladder cancer cell, a pancreatic cancer cell, a
CC   melanoma cell or a leukaemia cell) that expresses (I). Oligonucleotides
CC   hybridising to (II) are useful as diagnostic probes, antisense
CC   oligonucleotide probes or for encoding fragments of full length TAT
CC   polypeptide. (II) is also useful in chromosome and gene mapping and in
CC   the generation of antisense RNA and DNA probes, for constructing
CC   hybridisation probes for mapping the gene encoding TAT and for genetic
CC   analysis of individuals with genetic disorders. (II) is also useful for
CC   generating either transgenic animals or knockout animals, and in gene
CC   therapy. The TAT polypeptides and nucleic acids may also be used for
CC   tissue typing and the TAT polypeptides are useful for screening
CC   compounds that mimic the TAT polypeptide (agonist) or prevent the
CC   effect of TAT polypeptide (antagonist). The antibody is useful for
CC   staging TAT polypeptide-expressing cancers, purifying or
CC   immunoprecipitating TAT polypeptide from cells, for detection and
CC   quantitation of TAT polypeptide in vitro, e.g., in an enzyme linked
CC   immunosorbent assay (ELISA) or Western blot. The antibodies are also
CC   useful for treating a TAT-expressing cancer or alleviating one or more
CC   symptoms of cancer in a mammal. The present sequence represents the
CC   amino acid sequence of TAT138.
XX
SQ   Sequence   454 AA;

  Query Match            100.0%;  Score 2351;  DB 23;  Length 454;
  Best Local Similarity  100.0%;  Pred. No. 1.3e-244;
  Matches  454;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MESISMMGSPKSLSETCLPNGINGIKDARKVTVGVIGSGDFAKSLTIRLIRCGYHVVIGS 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MESISMMGSPKSLSETCLPNGINGIKDARKVTVGVIGSGDFAKSLTIRLIRCGYHVVIGS 6Q

Qy      61 RNPKFASEFFPHVVDVTHHEDALTKTNIIFVAIHREHYTSLWDLRHLLVGKILIDVSNNM 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 RNPKFASEFFPHVVDVTHHEDALTKTNIIFVAIHREHYTSLWDLRHLLVGKILIDVSNNM 120

Qy     121 RINQYPESNAEYLASLFPDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIE 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 RINQYPESNAEYLASLFPDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIE 180

Qy     181 LARQLNFIPIDLGSLSSAREIENLPLRLFTLWRGPVVVAISLATFFFLYSFVRDVIHPYA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 LARQLNFIPIDLGSLSSAREIENLPLRLFTLWRGPVVVAISLATFFFLYSFVRDVIHPYA 240

Qy     241 RNQQSDFYKIPIEIVNKTLPIVAITLLSLVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQ 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 RNQQSDFYKIPIEIVNKTLPIVAITLLSLVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQ 300

Qy     301 CRKQLGLLSFFFAMVHVAYSLCLPMRRSERYLFLNMAYQQVHANIENSWNEEEVWRIEMY 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CRKQLGLLSFFFAMVHVAYSLCLPMRRSERYLFLNMAYQQVHANIENSWNEEEVWRIEMY 360

Qy     361 ISFGIMSLGLLSLLAVTSIPSVSNALNWREFSFIQSTLGYVALLISTFHVLIYGWKRAFE 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 ISFGIMSLGLLSLLAVTSIPSVSNALNWREFSFIQSTLGYVALLISTFHVLIYGWKRAFE 420

Qy     421 EEYYRFYTPPNFVLALVLPSIVILDLLQLCRYPD 454
           |||||||||||||||||||||||||||||||||
Db     421 EEYYRFYTPPNFVLALVLPSIVILDLLQLCRYPD 454


RESULT 6
AAU10187
ID   AAU10187 standard; Protein; 490 AA.
XX
AC   AAU10187;
XX
DT   16-JAN-2002  (first entry)
XX
DE   Human Six-Transmembrane Protein of Prostate 1, STMP1.
XX
KW   Human; Six-Transmembrane Protein of Prostate 1; STM;1; prostate cancer;
KW   benign prostatic hyperplasia; acute prostatitis; testicular cancer;
KW   cryptorchidism; testicular disorder; proliferative disorder; lymphoma;
KW   leukaemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer;
KW   liver cancer; lung cancer; cytostatic.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Domain          209..230
FT                   /label= Transmembrane_domain_1
FT   Domain          255..273
FT                   /label= Transmembrane_domain_2
FT   Domain          304..325
FT                   /label= Transmembrane_domain_3
FT   Domain          360..380
FT                   /label= Transmembrane_domain_4
FT   Domain          393..415
FT                   /label= Transmembrane_domain_5
FT   Domain          432..452
FT                   /label= Transmembrane_domain_6
XX
PN   WO200172962-A2.
XX
PD   04-OCT-2001.
XX
PF   23-MAR-2001; 2001WO-US09410.
XX
PR   24-MAR-2000; 2000US-191929P.
XX
PA   (SAAT/) SAATCIOGLU F.
XX
```

PI Saatcioglu F;
XX
DR WPI; 2001-662926/76.
DR N-PSDB; AAS15793, AAS15801, AAS15802.
XX
PT New polynucleotide for the diagnosis, prevention and treatment for
PT prostate and testis disorders, particularly prostate cancer, comprises
PT prostate-specific or testis-specific nucleic acids -
XX
PS Claim 1; Fig 4B; 114pp; English.
XX
CC The invention relates to substantially pure prostate-specific or
CC testis-specific polypeptides and the nucleic acids encoding them.
CC Also included are vectors and host cells expressing the proteins, a
CC transgenic animal expressing the protein, antibodies against the
CC proteins, probes for detecting the nucleic acids, antisense molecules
CC for the nucleic acids and methods of isolating modulators of the
CC proteins. Compounds that modulate the prostate specific or testis
CC specific polypeptide are useful to diagnose, prevent or treat disorders
CC of the testis or prostate particularly prostate cancer, benign
CC prostatic hyperplasia, acute prostatitis, testicular cancer,
CC cryptorchidism, undescended, retractile, ascending or vanished
CC testis. Other proliferative disorders for which the modulators may be
CC used include lymphoma, leukaemia, melanoma, ovarian cancer, breast
CC cancer, pancreatic cancer, liver cancer and lung cancer. The
CC present sequence represents a prostate specific protein, Six-
CC Transmembrane Protein of Prostate 1, STM;1.
XX
SQ Sequence 490 AA;

Query Match 97.4%; Score 2290; DB 22; Length 490;
Best Local Similarity 99.8%; Pred. No. 5.5e-238;
Matches 443; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy       1 MESISMMGSPKSLSETCLPNGINGIKDARKVTVGVIGSGDFAKSLTIRLIRCGYHVVIGS 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MESISMMGSPKSLSETCLPNGINGIKDARKVTVGVIGSGDFAKSLTIRLIRCGYHVVIGS 60

Qy      61 RNPKFASEFFPHVVDVTHHEDALTKTNIIFVAIHREHYTSLWDLRHLLVGKILIDVSNNM 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 RNPKFASEFFPHVVDVTHHEDALTKTNIIFVAIHREHYTSLWDLRHLLVGKILIDVSNNM 120

Qy     121 RINQYPESNAEYLASLFPDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIE 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 RINQYPESNAEYLASLFPDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIE 180

Qy     181 LARQLNFIPIDLGSLSSAREIENLPLRLFTLWRGPVVVAISLATFFFLYSFVRDVIHPYA 240
          |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db     181 LARQLNFIPIDLGSLSSAREIENLPLRLFTFWRGPVVVAISLATFFFLYSFVRDVIHPYA 240

Qy     241 RNQQSDFYKIPIEIVNKTLPIVAITLLSLVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQ 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 RNQQSDFYKIPIEIVNKTLPIVAITLLSLVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQ 300

Qy     301 CRKQLGLLSFFFAMVHVAYSLCLPMRRSERYLFLNMAYQQVHANIENSWNEEEVWRIEMY 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CRKQLGLLSFFFAMVHVAYSLCLPMRRSERYLFLNMAYQQVHANIENSWNEEEVWRIEMY 360

Qy     361 ISFGIMSLGLLSLLAVTSIPSVSNALNWREFSFIQSTLGYVALLISTFHVLIYGWKRAFE 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 ISFGIMSLGLLSLLAVTSIPSVSNALNWREFSFIQSTLGYVALLISTFHVLIYGWKRAFE 420

Qy     421 EEYYRFYTPPNFVLALVLPSIVIL 444
          ||||||||||||||||||||||||
Db     421 EEYYRFYTPPNFVLALVLPSIVIL 444
```

RESULT 7
AAU10189
ID AAU10189 standard; Protein; 419 AA.
XX
AC AAU10189;
XX
DT 16-JAN-2002 (first entry)
XX
DE Human ORF3 of Six-Transmembrane Protein of Prostate 1.
XX
KW Human; Six-Transmembrane Protein of Prostate 1; STM;1; prostate cancer;
KW benign prostatic hyperplasia; acute prostatitis; testicular cancer;
KW cryptorchidism; testicular disorder; proliferative disorder; lymphoma;
KW leukaemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer;
KW liver cancer; lung cancer; cytostatic; ORF3.
XX
OS Homo sapiens.
XX
PN WO200172962-A2.
XX
PD 04-OCT-2001.
XX
PF 23-MAR-2001; 2001WO-US09410.
XX
PR 24-MAR-2000; 2000US-191929P.
XX
PA (SAAT/) SAATCIOGLU F.
XX
PI Saatcioglu F;
XX
DR WPI; 2001-662926/76.
DR N-PSDB; AAS15813.
XX
PT New polynucleotide for the diagnosis, prevention and treatment for
PT prostate and testis disorders, particularly prostate cancer, comprises
PT prostate-specific or testis-specific nucleic acids -
XX
PS Claim 1; Fig 4K; 114pp; English.
XX
CC The invention relates to substantially pure prostate-specific or
CC testis-specific polypeptides and the nucleic acids encoding them.
CC Also included are vectors and host cells expressing the proteins, a
CC transgenic animal expressing the protein, antibodies against the
CC proteins, probes for detecting the nucleic acids, antisense molecules
CC for the nucleic acids and methods of isolating modulators of the
CC proteins. Compounds that modulate the prostate specific or testis
CC specific polypeptide are useful to diagnose, prevent or treat disorders
CC of the testis or prostate particularly prostate cancer, benign
CC prostatic hyperplasia, acute prostatitis, testicular cancer,
CC cryptorchidism, undescended, retractile, ascending or vanished
CC testis. Other proliferative disorders for which the modulators may be
CC used include lymphoma, leukaemia, melanoma, ovarian cancer, breast
CC cancer, pancreatic cancer, liver cancer and lung cancer. The
CC present sequence is prostate specific protein, Six-Transmembrane
CC Protein of Prostate 1, STM;1, ORF3.
XX
SQ Sequence 419 AA;

Query Match 86.6%; Score 2036; DB 22; Length 419;
Best Local Similarity 99.7%; Pred. No. 1.1e-210;
Matches 394; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy       1 MESISMMGSPKSLSETCLPNGINGIKDARKVTVGVIGSGDFAKSLTIRLIRCGYHVVIGS 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MESISMMGSPKSLSETCLPNGINGIKDARKVTVGVIGSGDFAKSLTIRLIRCGYHVVIGS 60

Qy      61 RNPKFASEFFPHVVDVTHHEDALTKTNIIFVAIHREHYTSLWDLRHLLVGKILIDVSNNM 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 RNPKFASEFFPHVVDVTHHEDALTKTNIIFVAIHREHYTSLWDLRHLLVGKILIDVSNNM 120

Qy     121 RINQYPESNAEYLASLFPDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIE 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 RINQYPESNAEYLASLFPDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIE 180

Qy     181 LARQLNFIPIDLGSLSSAREIENLPLRLFTLWRGPVVVAISLATFFFLYSFVRDVIHPYA 240
          |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db     181 LARQLNFIPIDLGSLSSAREIENLPLRLFTFWRGPVVVAISLATFFFLYSFVRDVIHPYA 240

Qy     241 RNQQSDFYKIPIEIVNKTLPIVAITLLSLVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQ 300
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 RNQQSDFYKIPIEIVNKTLPIVAITLLSLVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQ 300

Qy    301 CRKQLGLLSFFFAMVHVAYSLCLPMRRSERYLFLNMAYQQVHANIENSWNEEEVWRIEMY 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CRKQLGLLSFFFAMVHVAYSLCLPMRRSERYLFLNMAYQQVHANIENSWNEEEVWRIEMY 360

Qy    361 ISFGIMSLGLLSLLAVTSIPSVSNALNWREFSFIQ 395
          |||||||||||||||||||||||||||||||||||
Db    361 ISFGIMSLGLLSLLAVTSIPSVSNALNWREFSFIQ 395


RESULT 8
ABG12306
ID   ABG12306 standard; Protein; 576 AA.
XX
AC   ABG12306;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #12297.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS76493.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 42665; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   576 AA;

  Query Match           83.9%;  Score 1972.5;  DB 22;  Length 576;
  Best Local Similarity  87.4%;  Pred. No. 1.3e-203;
  Matches  390;  Conservative    0;  Mismatches    1;  Indels   55;  Gaps    1;

Qy      1 MESISMMGSPKSLSETCLPNGINGIKDARKVTVGVIGSGDFAKSLTIRLIRCGYHVVIGS 60
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db      1 MESISMMGSPKSLSETFLPNGINGIKDARKVTVGVIGSGDFAKSLTIRLIRCGYHVVIGS 60

Qy     61 RNPKFASEFFPHVVDVTHHEDALTKTNIIFVAIHREHYTSLWDLRHLLVGKILIDVSNNM 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 RNPKFASEFFPHVVDVTHHEDALTKTNIIFVAIHREHYTSLWDLRHLLVGKILIDVSNNM 120

Qy    121 RINQYPESNAEYLASLFPDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIE 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 RINQYPESNAEYLASLFPDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIE 180

Qy    181 LARQLNFIPIDLGSLSSAREIENLPLRLFTLWRGPVVVAISLATFFFLYSFVRDVIHPYA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 LARQLNFIPIDLGSLSSAREIENLPLRLFTLWRGPVVVAISLATFFFLYSFVRDVIHPYA 240

Qy    241 RNQQSDFYKIPIEIVNKTLPIVAITLLSLVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQ 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 RNQQSDFYKIPIEIVNKTLPIVAITLLSLVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQ 300

Qy    301 CRKQLGLLSFFFAMVHVAYSLCLPMRRSERYLFLNMAYQQVHANIENSWNEEEVWRIEMY 360
          |||||||||||||||||||||||||||||||||||
Db    301 CRKQLGLLSFFFAMVHVAYSLCLPMRRSERYLFLNMAYQ---------------------- 339

Qy    361 ISFGIMSLGLLSLLAVTSIPSVSNALNWREFSFIQSTLGYVALLISTFHVLIYGWKRAFE 420
                                              ||||||||||||||||||||||||
Db    340 ------------------------------------QSTLGYVALLISTFHVLIYGWKRAFE 365

Qy    421 EEYYRFYTPPNFVLALVLPSIVILDL 446
          ||||||||||||||||||||||||||
Db    366 EEYYRFYTPPNFVLALVLPSIVILDL 391


RESULT 9
ABG00113
ID   ABG00113 standard; Protein; 1273 AA.
XX
AC   ABG00113;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #104.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS64300.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
```

```
XX
PS   Claim 20; SEQ ID No 30472; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1273 AA;

  Query Match             83.9%;  Score 1972.5;  DB 22;  Length 1273;
  Best Local Similarity   87.4%;  Pred. No. 4.5e-203;
  Matches  390;  Conservative    0;  Mismatches    1;  Indels   55;  Gaps    1;

Qy       1 MESISMMGSPKSLSETCLPNGINGIKDARKVTVGVIGSGDFAKSLTIRLIRCGYHVVIGS 60
           |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db      73 MESISMMGSPKSLSETFLPNGINGIKDARKVTVGVIGSGDFAKSLTIRLIRCGYHVVIGS 132

Qy      61 RNPKFASEFFPHVVDVTHHEDALTKTNIIFVAIHREHYTSLWDLRHLLVGKILIDVSNNM 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     133 RNPKFASEFFPHVVDVTHHEDALTKTNIIFVAIHREHYTSLWDLRHLLVGKILIDVSNNM 192

Qy     121 RINQYPESNAEYLASLFPDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIE 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     193 RINQYPESNAEYLASLFPDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIE 252

Qy     181 LARQLNFIPIDLGSLSSAREIENLPLRLFTLWRGPVVVAISLATFFFLYSFVRDVIHPYA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     253 LARQLNFIPIDLGSLSSAREIENLPLRLFTLWRGPVVVAISLATFFFLYSFVRDVIHPYA 312

Qy     241 RNQQSDFYKIPIEIVNKTLPIVAITLLSLVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQ 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     313 RNQQSDFYKIPIEIVNKTLPIVAITLLSLVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQ 372

Qy     301 CRKQLGLLSFFFAMVHVAYSLCLPMRRSERYLFLNMAYQQVHANIENSWNEEEVWRIEMY 360
           ||||||||||||||||||||||||||||||||||||
Db     373 CRKQLGLLSFFFAMVHVAYSLCLPMRRSERYLFLNMAYQ---------------------- 411

Qy     361 ISFGIMSLGLLSLLAVTSIPSVSNALNWREFSFIQSTLGYVALLISTFHVLIYGWKRAFE 420
                                           ||||||||||||||||||||||||||
Db     412 -----------------------------------QSTLGYVALLISTFHVLIYGWKRAFE 437

Qy     421 EEYYRFYTPPNFVLALVLPSIVILDL 446
           ||||||||||||||||||||||||||
Db     438 EEYYRFYTPPNFVLALVLPSIVILDL 463


RESULT 10
AAE02841
ID   AAE02841 standard; Protein; 450 AA.
XX
AC   AAE02841;
XX
DT   06-AUG-2001  (first entry)
XX
DE   Human STEAP-2 protein, alternative version.
XX
KW   Human; cytostatic; antiproliferative; vaccine; gene therapy;
KW   six transmembrane epithelial antigen of the prostate-2; STEAP-2;
KW   chromosome 7q21; cancer; prostate; colon; bladder; lung; ovarian;
KW   pancreatic.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 335..336
FT                   /note= "Encoded by GACTGAGCT"
FT   Misc-difference 375..376
FT                   /note= "Encoded by AGATGAAGT"
FT   Misc-difference 415..416
FT                   /note= "Encoded by GAGTAAGGC"
FT   Misc-difference 445..446
FT                   /note= "Encoded by ACATGAAGT"
FT   Misc-difference 448..449
FT                   /note= "Encoded by AATTAATTC"
XX
PN   WO200140276-A2.
XX
PD   07-JUN-2001.
XX
PF   06-DEC-2000; 2000WO-US33040.
XX
PR   06-DEC-1999;   99US-0455486.
XX
PA   (UROG-) UROGENESYS INC.
XX
PI   Afar DEH,  Hubert RS,  Raitano AB,  Saffran DC,  Mitchell SC,  Faris M;
PI   Jakobovits A;
XX
DR   WPI; 2001-367804/38.
DR   N-PSDB; AAD07072.
XX
PT   New STEAP (six transmembrane epithelial antigen of the prostate)
PT   proteins, expressed in human cancers, useful for detecting and treating
PT   cancer -
XX
PS   Claim 1; Page 175-176; 187pp; English.
XX
CC   The present sequence is an alternative version of human six transmembrane
CC   epithelial antigen of the prostate (STEAP)-2 protein. STEAP is a member
CC   of cell surface serpentine transmembrane antigens. STEAP-2 gene is
CC   located on chromosome 7q21 and is used in gene therapy. Inhibiting the
CC   development or progression of a cancer (eg. prostate, colon, bladder,
CC   lung, ovarian and pancreatic) expressing STEAP or inhibiting growth or
CC   killing cells expressing STEAP in a patient, comprises administering a
CC   vaccine composition to the patient. Treating a patient with a cancer that
CC   expresses STEAP, or inhibiting growth or killing cells expressing STEAP,
CC   comprises administering to the patient a vector encoding single chain
CC   monoclonal antibody that comprises the variable domains of the heavy and
CC   light chains of the monoclonal antibody that specifically binds to STEAP,
CC   such that the vector delivers the single chain monoclonal antibody coding
CC   sequence to the cancer cells and the encoded single chain monoclonal
CC   antibody is expressed intracellularly.
CC   Note: This sequence is stated to be the same as that being shown as
CC   SEQ ID NO:8 (AAE02781) in figure 9A-9C of the specification. However
CC   the present sequence lacks several amino acids at its N-terminal end
CC   and has additional amino acids at its C-terminal end when compared with
CC   the sequence shown in figure 9A-9C.
XX
SQ   Sequence   450 AA;

  Query Match             73.9%;  Score 1738;  DB 22;  Length 450;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-178;
  Matches  336;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     119 NMRINQYPESNAEYLASLFPDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQV 178
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 NMRINQYPESNAEYLASLFPDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQV 60
```

```
Qy    179 IELARQLNFIPIDLGSLSSAREIENLPLRLFTLWRGPVVVAISLATFFFLYSFVRDVIHP 238
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 IELARQLNFIPIDLGSLSSAREIENLPLRLFTLWRGPVVVAISLATFFFLYSFVRDVIHP 120

Qy    239 YARNQQSDFYKIPIEIVNKTLPIVAITLLSLVYLAGLLAAAYQLYYGTKYRRFPPWLETW 298
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 YARNQQSDFYKIPIEIVNKTLPIVAITLLSLVYLAGLLAAAYQLYYGTKYRRFPPWLETW 180

Qy    299 LQCRKQLGLLSFFFAMVHVAYSLCLPMRRSERYLFLNMAYQQVHANIENSWNEEEVWRIE 358
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 LQCRKQLGLLSFFFAMVHVAYSLCLPMRRSERYLFLNMAYQQVHANIENSWNEEEVWRIE 240

Qy    359 MYISFGIMSLGLLSLLAVTSIPSVSNALNWREFSFIQSTLGYVALLISTFHVLIYGWKRA 418
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 MYISFGIMSLGLLSLLAVTSIPSVSNALNWREFSFIQSTLGYVALLISTFHVLIYGWKRA 300

Qy    419 FEEEYYRFYTPPNFVLALVLPSIVILDLLQLCRYPD 454
          ||||||||||||||||||||||||||||||||||||
Db    301 FEEEYYRFYTPPNFVLALVLPSIVILDLLQLCRYPD 336
```

RESULT 11
AAB49483
ID AAB49483 standard; Protein; 488 AA.
XX
AC AAB49483;
XX
DT 08-MAR-2001 (first entry)
XX
DE Rat p-HYDE.
XX
KW Rat; p-HYDE; cytostatic; gene therapy; apoptosis; leukaemia; prostate;
KW tumour suppressor gene; DNA repair; cancer; melanoma; lymphoma;
KW colorectal; pancreatic; breast; brain; gastric carcinoma.
XX
OS Rattus sp.
XX
PN WO200071564-A2.
XX
PD 30-NOV-2000.
XX
PF 01-MAY-2000; 2000WO-US11456.
XX
PR 29-APR-1999; 99US-0131607.
PR 29-APR-1999; 99US-0302457.
PR 26-NOV-1999; 99US-0499817.
XX
PA (UYTE-) UNIV TENNESSEE RES CORP.
XX
PI Steiner MS, Wang C, Rinaldy A, Menon R;
XX
DR WPI; 2001-032016/04.
DR N-PSDB; AAC89167.
XX
PT New isolated nucleic acid encoding a mammalian p-Hyde protein of the
PT p-Hyde family is useful for treating cancer, e.g. prostate cancer -
XX
PS Claim 31; Page 20; 171pp; English.
XX
CC The present protein is rat p-HYDE. p-HYDE induces susceptibility of a
CC cancer cell to cell death. The p-HYDE gene is associated with the
CC regression of tumour growth in vivo, the induction to susceptibility to
CC apoptosis caused by UV or chemotherapy induced DNA damage and prevention
CC of DNA repair with the upregulation of apoptosis as the result of UV
CC damage and the failure to repair DNA. The present sequence may be used to
CC treat cancer, preferably melanoma, lymphoma, leukaemia, prostate,
CC colorectal, pancreatic, breast, brain or gastric carcinoma.
XX
SQ Sequence 488 AA;

Query Match 56.0%; Score 1316; DB 22; Length 488;
Best Local Similarity 56.2%; Pred. No. 7.6e-133;
Matches 248; Conservative 81; Mismatches 100; Indels 12; Gaps 4;

```
Qy      6 MMGSPKSLSETCLPNGINGIKDARKVTVGVIGSGDFAKSLTIRLIRCGYHVVIGSRNPKF 65
          :: |  ||:|  :|      |:| |  ||::||||||:||  ||:  |: ||:||||||
Db     14 LVDSDGSLAE--VP------KEAPK--VGILGSGDFARSLATRLVGSGFFVVVGSRNPKR 63

Qy     66 ASEFFPHVVDVTHHEDALTKTNIIFVAIHREHYTSLWDLRHLLVGKILIDVSNNMRIN-- 123
            :  || :  ||  |:|::   :||||: ||||:||  |   | ||||:||||
Db     64 TAGLFPSLAQVTFQEEAVSSPEVIFVAVFREHYSSLCSLADQLAGKILVDVSNPTEKERL 123

Qy    124 QYPESNAEYLASLFPDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIELAR 183
          |: :||||||||||   :|| |||:|||||| ||:| :||| || : ::|:  | |:||
Db    124 QHRQSNAEYLASLFPACTVVKAFNVISAWALQAGPRDGNRQVLICGDQLEAKHTVSEMAR 183

Qy    184 QLNFIPIDLGSLSSAREIENLPLRLFTLWRGPVVVAISLATFFFLYSFVRDVIHPYARNQ 243
            : | |:|:|||:||||:| :||||   |: | ::|: |:|  : |:|:|||: || |
Db    184 AMGFTPLDMGSLASAREVEAIPLRLLPSWKVPTLLALGLSTQSYAYNFIRDVLQPYIRKD 243

Qy    244 QSDFYKIPIEIVNKTLPIVAITLLSLVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQCRK 303
          :: |||:|: :|| |:| ||  ||||||| |:|||| ||  ||||:||| ||: ||| ||
Db    244 ENKFYKMPLSVVNTTIPCVAYVLLSLVYLPGVLAAALQLRRGTKYQRFPDWLDHWLQHRK 303

Qy    304 QLGLLSFFFAMVHVAYSLCLPMRRSERYLFLNMAYQQVHANIENSWNEEEVWRIEMYISF 363
          |:|||||||||:|  || |||:||| ||  :|:| :|| ||    | ||||||:|:|:|
Db    304 QIGLLSFFFAMLHALYSFCLPLRRSHRYDLVNLAVKQVLANKSRLWVEEEVWRMEIYLSL 363

Qy    364 GIMSLGLLSLLAVTSIPSVSNALNWREFSFIQSTLGYVALLISTFHVLIYGWKRAFEEEY 423
          |:::||:|||||||||||::|:|||:||||:|||||:|||::|| | | ||| ||||| :
Db    364 GVLALGMLSLLAVTSIPSIANSLNWKEFSFVQSTLGFVALMLSTMHTLTYGWTRAFEENH 423

Qy    424 YRFYTPPNFVLALVLPSIVIL 444
          |:|| || | | |:|| ::||
Db    424 YKFYLPPTFTLTLLLPCVIIL 444
```

RESULT 12
ABB83365
ID ABB83365 standard; Protein; 526 AA.
XX
AC ABB83365;
XX
DT 02-SEP-2002 (first entry)
XX
DE Murine Tumour Suppressor Activated Pathway 6, TSAP6.
XX
KW Neuroprotective; cytostatic; TSAP6; apoptosis; cancer; neurodegeneration;
KW Tumour Suppressor Activated Pathway 6; TSAP6 binding protein; tumour;
KW murine.
XX
OS Mus musculus.
XX
PN WO200252274-A2.
XX
PD 04-JUL-2002.
XX
PF 24-DEC-2001; 2001WO-FR04188.
XX
PR 26-DEC-2000; 2000FR-0017027.
PR 18-SEP-2001; 2001WO-FR02896.
XX
PA (MOLE-) MOLECULAR ENGINES LAB.
XX
PI Amson R, Telerman A, Passer B;
XX
DR WPI; 2002-508914/54.
DR N-PSDB; ABN85026.
XX
PT Identifying compounds that inhibit binding of TSAP6 to proteins for
PT treating cancer and neurodegeneration -
XX
PS Disclosure; Page 75-77; 79pp; French.
XX
CC The present invention relates to a method for identifying a compound
CC (I) that inhibits binding of TSAP (Tumour Suppressor Activated Pathway) 6

```
CC   to a TSAP6 binding protein. The present sequence is murine TSAP6, which
CC   was used in the method of the invention. (I) are useful for modulating
CC   tumour reversal and/or apoptosis for treating cancer or
CC   neurodegeneration.
XX
SQ   Sequence   526 AA;

  Query Match             56.0%;  Score 1316;  DB 23;  Length 526;
  Best Local Similarity   56.2%;  Pred. No. 8.5e-133;
  Matches  248;  Conservative   81;  Mismatches  100;  Indels   12;  Gaps    4;

Qy       6 MMGSPKSLSETCLPNGINGIKDARKVTVGVIGSGDFAKSLTIRLIRCGYHVVIGSRNPKF 65
Db      52 LVDSDGSLAE--VP------KEAPK--VGILGSGDFARSLATRLVGSGFSVVVGSRNPKR 101

Qy      66 ASEFFPHVVDVTHHEDALTKTNIIFVAIHREHYTSLWDLRHLLVGKILIDVSNNMRIN-- 123
Db     102 TAGLFPSLAQVTFQEEAVSSPEVIFVAVFREHYSSLCSLADQLAGKILVDVSNPTEKEHL 161

Qy     124 QYPESNAEYLASLFPDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIELAR 183
Db     162 QHRQSNAEYLASLFPACTVVKAFNVISAWALQAGPRDGNRQVLICSDQPEAKRTISEMAR 221

Qy     184 QLNFIPIDLGSLSSAREIENLPLRLFTLWRGPVVVAISLATFFFLYSFVRDVIHPYARNQ 243
Db     222 AMGFTPLDMGSLASAREVEAIPLRLLPSWKVPTLLALGLFVCFYTYNFIRDVLQPYIRKD 281

Qy     244 QSDFYKIPIEIVNKTLPIVAITLLSLVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQCRK 303
Db     282 ENKFYKMPLSVVNTTLPCVAYVLLSLVYLPGVLAAALQLRRGTKYQRFPDWLDHWLQHRK 341

Qy     304 QLGLLSFFFAMVHVAYSLCLPMRRSERYLFLNMAYQQVHANIENSWNEEEVWRIEMYISF 363
Db     342 QIGLLSFFFAMLHALYSFCLPLRRSHRYDLVNLAVKQVLANKSRLWVEEEVWRMEIYLSL 401

Qy     364 GIMSLGLLSLLAVTSIPSVSNALNWREFSFIQSTLGYVALLISTFHVLIYGWKRAFEEEY 423
Db     402 GVLALGMLSLLAVTSLPSIANSLNWKEFSFVQSTLGFVALILSTMHTLTYGWTRAFEENH 461

Qy     424 YRFYTPPNFVLALVLPSIVIL 444
Db     462 YKFYLPPTFTLTLLLPCVIIL 482


RESULT 13
AAB85775
ID   AAB85775 standard; Protein; 488 AA.
XX
AC   AAB85775;
XX
DT   29-OCT-2001  (first entry)
XX
DE   Human drug metabolizing enzyme (ID No. 5629033CD1).
XX
KW   Drug metabolizing enzyme; DME; immunosuppressive; cytostatic; opthalmic;
KW   hepatotrophic; antiallergic; antiasthmatic; antibacterial; antiviral;
KW   antisense therapy; gene therapy; human.
XX
OS   Homo sapiens.
XX
PN   WO200159127-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-US04423.
XX
PR   11-FEB-2000; 2000US-0181856.
PR   17-FEB-2000; 2000US-0183684.
PR   25-FEB-2000; 2000US-0185141.
PR   03-MAR-2000; 2000US-0186818.
PR   09-MAR-2000; 2000US-0188345.
PR   17-MAR-2000; 2000US-0189997.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Tang YT,  Yue H,  Baughn MR,  Yao MG,  Bandman O,  Azimzai Y,  Lal P;
PI   Gandhi AR,  Ring HZ,  Shih LL,  Yang J,  Policky JL;
XX
DR   WPI; 2001-514673/56.
DR   N-PSDB; AAH76202.
XX
PT   Isolated polypeptide encoding a drug metabolizing enzyme useful for the
PT   diagnosis, treatment, and prevention of autoimmune/inflammatory, cell
PT   proliferative, developmental and endocrine disorders -
XX
PS   Claim 1; Page 134-135; 150pp; English.
XX
CC   The invention provides human drug metabolizing enzymes (DME) and
CC   polynucleotides encoding the DMEs. The DME can be expressed by standard
CC   recombinant methodology. DMEs and their agonists and antagonists are
CC   useful for the diagnosis, treatment, and prevention of autoimmune/
CC   inflammatory, cell proliferative, developmental, endocrine such as
CC   aneurysm, eye, metabolic, and gastrointestinal disorders, including liver
CC   disorders and infection. The present sequence represents a human DME.
XX
SQ   Sequence   488 AA;

  Query Match             54.1%;  Score 1272;  DB 22;  Length 488;
  Best Local Similarity   54.5%;  Pred. No. 4.3e-128;
  Matches  242;  Conservative   82;  Mismatches  108;  Indels   12;  Gaps    4;

Qy       3 SISMMGSPKSLSETCLPNGINGIKDARKVTVGVIGSGDFAKSLTIRLIRCGYHVVIGSRN 62
Db      11 SLHLVDSDSSLAK--VPD------EAPK--VGILGSGDFARSLATRLVGSGFKVVVGSRN 60

Qy      63 PKFASEFFPHVVDVTHHEDALTKTNIIFVAIHREHYTSLWDLRHLLVGKILIDVSN--NM 120
Db      61 PKRTARLFPSAAQVTFQEEAVSSPEVIFVAVFREHYSSLCSLSDQLAGKILVDVSNPTEQ 120

Qy     121 RINQYPESNAEYLASLFPDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIE 180
Db     121 EHLQHRESNAEYLASLFPTCTVVKAFNVISAWTLQAGPRDGNRQVPICGDQPEAKRAVSE 180

Qy     181 LARQLNFIPIDLGSLSSAREIENLPLRLFTLWRGPVVVAISLATFFFLYSFVRDVIHPYA 240
Db     181 MALAMGFMPVDMGSLASAWEVEAMPLRLLPAWKVPTLLALGLFVCFYAYNFVRDVLQPYV 240

Qy     241 RNQQSDFYKIPIEIVNKTLPIVAITLLSLVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQ 300
Db     241 QESQNKFFKLPVSVVNTTLPCVAYVLLSLVYLPGVLAAALQLRRGTKYQRFPDWLDHWLQ 300

Qy     301 CRKQLGLLSFFFAMVHVAYSLCLPMRRSERYLFLNMAYQQVHANIENSWNEEEVWRIEMY 360
Db     301 HRKQIGLLSFFCAALHALYSFCLPLRRAHRYDLVNLAVKQVLANKSHLWVEEEVWRMEIY 360

Qy     361 ISFGIMSLGLLSLLAVTSIPSVSNALNWREFSFIQSTLGYVALLISTFHVLIYGWKRAFE 420
Db     361 LSLGVLALGTLSLLAVTSLPSIANSLNWREFSFVQSSLGFVALVLSTLHTLTYGWTRAFE 420

Qy     421 EEYYRFYTPPNFVLALVLPSIVIL 444
Db     421 ESRYKFYLPPTFTLTLLVPCVVIL 444


RESULT 14
ABB83366
ID   ABB83366 standard; Protein; 488 AA.
XX
AC   ABB83366;
XX
DT   02-SEP-2002  (first entry)
XX
DE   Human Tumour Suppressor Activated Pathway 6, TSAP6.
XX
KW   Neuroprotective; cytostatic; TSAP6; apoptosis; cancer; neurodegeneration;
KW   Tumour Suppressor Activated Pathway 6; TSAP6 binding protein; tumour;
```

```
KW   human.
XX
OS   Homo sapiens.
XX
PN   WO200252274-A2.
XX
PD   04-JUL-2002.
XX
PF   24-DEC-2001; 2001WO-FR04188.
XX
PR   26-DEC-2000; 2000FR-0017027.
PR   18-SEP-2001; 2001WO-FR02896.
XX
PA   (MOLE-) MOLECULAR ENGINES LAB.
XX
PI   Amson R,  Telerman A,  Passer B;
XX
DR   WPI; 2002-508914/54.
DR   N-PSDB; ABN85027.
XX
PT   Identifying compounds that inhibit binding of TSAP6 to proteins for
PT   treating cancer and neurodegeneration -
XX
PS   Disclosure; Page 78-79; 79pp; French.
XX
CC   The present invention relates to a method for identifying a compound
CC   (I) that inhibits binding of TSAP (Tumour Suppressor Activated Pathway) 6
CC   to a TSAP6 binding protein. The present sequence is human TSAP6, which
CC   was used in the method of the invention. (I) are useful for modulating
CC   tumour reversal and/or apoptosis for treating cancer or
CC   neurodegeneration.
XX
SQ   Sequence   488 AA;

  Query Match             54.1%;  Score 1272;  DB 23;  Length 488;
  Best Local Similarity   54.5%;  Pred. No. 4.3e-128;
  Matches  242;  Conservative   82;  Mismatches  108;  Indels   12;  Gaps    4;

Qy       3 SISMMGSPKSLSETCLPNGINGIKDARKVTVGVIGSGDFAKSLTIRLIRCGYHVVIGSRN 62
           |: :: |  ||::  :|:      :| |  ||::|||||:||  ||:  |: ||:||||
Db      11 SLHLVDSDSSLAK--VPD------EAPK--VGILGSGDFARSLATRLVGSGFKVVVGSRN 60

Qy      63 PKFASEFFPHVVDVTHHEDALTKTNIIFVAIHREHYTSLWDLRHLLVGKILIDVSN--NM 120
           ||  :  ||    ||  |:|::   :||||: ||||:||  |   | ||||:||||
Db      61 PKRTARLFPSAAQVTFQEEAVSSPEVIFVAVFREHYSSLCSLSDQLAGKILVDVSNPTEQ 120

Qy     121 RINQYPESNAEYLASLFPDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIE 180
             |: ||||||||||||   :|| |||:||| || ||:| :||| || :  :|:: | |
Db     121 EHLQHRESNAEYLASLFPTCTVVKAFNVISAWTLQAGPRDGNRQVPICGDQPEAKRAVSE 180

Qy     181 LARQLNFIPIDLGSLSSAREIENLPLRLFTLWRGPVVVAISLATFFFLYSFVRDVIHPYA 240
           :|  : |:|:|:|||:|| |:| :||||   |: | ::|: |   |: |:|||||: ||
Db     181 MALAMGFMPVDMGSLASAWEVEAMPLRLLPAWKVPTLLALGLFVCFYAYNFVRDVLQPYV 240

Qy     241 RNQQSDFYKIPIEIVNKTLPIVAITLLSLVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQ 300
           :  |: |:|:|: :|| ||| ||  ||||||| |:|||| ||  ||||:||| ||: |||
Db     241 QESQNKFFKLPVSVVNTTLPCVAYVLLSLVYLPGVLAAALQLRRGTKYQRFPDWLDHWLQ 300

Qy     301 CRKQLGLLSFFFAMVHVAYSLCLPMRRSERYLFLNMAYQQVHANIENSWNEEEVWRIEMY 360
            |||:|||||| | :|  || |||:||: ||  :|:| :|| ||  : | ||||||:|:|
Db     301 HRKQIGLLSFFCAALHALYSFCLPLRRAHRYDLVNLAVKQVLANKSHLWVEEEVWRMEIY 360

Qy     361 ISFGIMSLGLLSLLAVTSIPSVSNALNWREFSFIQSTLGYVALLISTFHVLIYGWKRAFE 420
           :| |:::|| |||||||:||::|:||||||||:||:||:|||::|| | | ||| ||||
Db     361 LSLGVLALGTLSLLAVTSLPSIANSLNWREFSFVQSSLGFVALVLSTLHTLTYGWTRAFE 420

Qy     421 EEYYRFYTPPNFVLALVLPSIVIL 444
           |  |:|| || | | |::| :|||
Db     421 ESRYKFYLPPTFTLTLLVPCVVIL 444


RESULT 15
AAU10220
```

```
ID   AAU10220 standard; Protein; 488 AA.
XX
AC   AAU10220;
XX
DT   16-JAN-2002  (first entry)
XX
DE   Human Six-Transmembrane Protein of Prostate 3, STMP3.
XX
KW   Human; Six-Transmembrane Protein of Prostate 3; STMP3; prostate cancer;
KW   benign prostatic hyperplasia; acute prostatitis; testicular cancer;
KW   cryptorchidism; testicular disorder; proliferative disorder; lymphoma;
KW   leukaemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer;
KW   liver cancer; lung cancer; cytostatic.
XX
OS   Homo sapiens.
XX
PN   WO200172962-A2.
XX
PD   04-OCT-2001.
XX
PF   23-MAR-2001; 2001WO-US09410.
XX
PR   24-MAR-2000; 2000US-191929P.
XX
PA   (SAAT/) SAATCIOGLU F.
XX
PI   Saatcioglu F;
XX
DR   WPI; 2001-662926/76.
DR   N-PSDB; AAS15815.
XX
PT   New polynucleotide for the diagnosis, prevention and treatment for
PT   prostate and testis disorders, particularly prostate cancer, comprises
PT   prostate-specific or testis-specific nucleic acids  -
XX
PS   Claim 1; Fig 4M; 114pp; English.
XX
CC   The invention relates to substantially pure prostate-specific or
CC   testis-specific polypeptides and the nucleic acids encoding them.
CC   Also included are vectors and host cells expressing the proteins, a
CC   transgenic animal expressing the protein, antibodies against the
CC   proteins, probes for detecting the nucleic acids, antisense molecules
CC   for the nucleic acids and methods of isolating modulators of the
CC   proteins. Compounds that modulate the prostate specific or testis
CC   specific polypeptide are useful to diagnose, prevent or treat disorders
CC   of the testis or prostate particularly prostate cancer, benign
CC   prostatic hyperplasia, acute prostatitis, testicular cancer,
CC   cryptorchidism, undescended, retractile, ascending or vanished
CC   testis. Other proliferative disorders for which the modulators may be
CC   used include lymphoma, leukaemia, melanoma, ovarian cancer, breast
CC   cancer, pancreatic cancer, liver cancer and lung cancer. The
CC   present sequence is a prostate specific protein, Six-Transmembrane
CC   Protein of Prostate 3, STMP3.
XX
SQ   Sequence   488 AA;

  Query Match             54.0%;  Score 1269;  DB 22;  Length 488;
  Best Local Similarity   54.3%;  Pred. No. 9e-128;
  Matches  241;  Conservative   83;  Mismatches  108;  Indels   12;  Gaps    4;

Qy       3 SISMMGSPKSLSETCLPNGINGIKDARKVTVGVIGSGDFAKSLTIRLIRCGYHVVIGSRN 62
           |: :: |  ||::  :|:      :| |  ||::|||||:||  ||:  |: ||:||||
Db      11 SLHLVDSDSSLAK--VPD------EAPK--VGILGSGDFARSLATRLVGSGFKVVVGSRN 60

Qy      63 PKFASEFFPHVVDVTHHEDALTKTNIIFVAIHREHYTSLWDLRHLLVGKILIDVSN--NM 120
           ||  :  :|    ||  |:|::   :||||: ||||:||  |   | ||||:||||
Db      61 PKRTARLYPSAAQVTFQEEAVSSPEVIFVAVFREHYSSLCSLSDQLAGKILVDVSNPTEQ 120

Qy     121 RINQYPESNAEYLASLFPDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIE 180
             |: ||||||||||||   :|| |||:||| || ||:| :||| || :  :|:: | |
Db     121 EHLQHRESNAEYLASLFPTCTVVKAFNVISAWTLQAGPRDGNRQVPICGDQPEAKRAVSE 180

Qy     181 LARQLNFIPIDLGSLSSAREIENLPLRLFTLWRGPVVVAISLATFFFLYSFVRDVIHPYA 240
```

```
          :|  : |:|:|:|||:|| |:| :|||| |: | ::|: |  |: |:|||||: ||
Db    181 MALAMGFMPVDMGSLASAWEVEAMPLRLLPAWKVPTLLALGLFVCFYAYNFVRDVLQPYV 240

Qy    241 RNQQSDFYKIPIEIVNKTLPIVAITLLSLVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQ 300
          :  |: |:|:|: :|| ||| ||  |||||| |:|||| ||  ||||:||| ||: |||
Db    241 QESQNKFFKLPVSVVNTTLPCVAYVLLSLVYLPGVLAAALQLRRGTKYQRFPDWLDHWLQ 300

Qy    301 CRKQLGLLSFFFAMVHVAYSLCLPMRRSERYLFLNMAYQQVHANIENSWNEEEVWRIEMY 360
           |||:|||||| | :|  || |||:||: ||  :|:| :|| ||  : | |||||||:|:|
Db    301 HRKQIGLLSFFCAALHALYSFCLPLRRAHRYDLVNLAVKQVLANKSHLWVEEEVWRMEIY 360

Qy    361 ISFGIMSLGLLSLLAVTSIPSVSNALNWREFSFIQSTLGYVALLISTFHVLIYGWKRAFE 420
          :| |:::|| |||||||||:||::|:|||||||||:||:||:|||::|| | | ||| ||||
Db    361 LSLGVLALGTLSLLAVTSLPSIANSLNWREFSFVQSSLGFVALVLSTLHTLTYGWTRAFE 420

Qy    421 EEYYRFYTPPNFVLALVLPSIVIL 444
          |  |:|| || | | |::| :|||
Db    421 ESRYKFYLPPTFTLTLLVPCVVIL 444
```

Search completed: January 17, 2003, 18:59:10
Job time : 46 secs

OM protein - protein search, using sw model

Run on: January 17, 2003, 18:58:14 ; Search time 18 Seconds
(without alignments)
742.111 Million cell updates/sec

Title: US-09-455-486-6
Perfect score: 2351
Sequence: 1 MESISMMGSPKSLSETCLPN.....ALVLPSIVILDLLQLCRYPD 454

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	901	38.3	173	4	US-09-323-873A-8	Sequence 8, Appli
2	736	31.3	141	3	US-09-083-521-1	Sequence 1, Appli
3	717	30.5	339	4	US-09-323-873A-2	Sequence 2, Appli
4	144.5	6.1	227	4	US-09-655-270A-15	Sequence 15, Appl
5	144.5	6.1	227	4	US-09-651-941-17	Sequence 17, Appl
6	144.5	6.1	227	4	US-09-955-597-17	Sequence 17, Appl
7	107.5	4.6	695	1	US-08-487-886-2	Sequence 2, Appli
8	107.5	4.6	695	3	US-08-482-855-2	Sequence 2, Appli
9	107.5	4.6	695	4	US-08-474-986-2	GENERAL INFORMA
10	107	4.6	34	4	US-09-323-873A-20	Sequence 20, Appl
11	101	4.3	476	4	US-09-316-083-3	Sequence 3, Appli
12	99	4.2	531	2	US-08-724-974A-2	Sequence 2, Appli
13	97.5	4.1	390	3	US-08-460-576-2	Sequence 2, Appli
14	97	4.1	692	4	US-07-757-342D-6	Sequence 6, Appli
15	93	4.0	365	2	US-08-724-974A-3	Sequence 3, Appli
16	91.5	3.9	495	4	US-08-985-343-2	Sequence 2, Appli
17	91	3.9	808	4	US-09-134-001C-3105	Sequence 3105, Ap
18	89.5	3.8	495	1	US-07-841-997A-2	Sequence 2, Appli
19	89.5	3.8	495	1	US-08-290-301-2	Sequence 2, Appli
20	89.5	3.8	495	2	US-08-588-983-2	Sequence 2, Appli
21	89.5	3.8	495	2	US-08-588-976-2	Sequence 2, Appli
22	89.5	3.8	495	4	US-09-013-598-2	Sequence 2, Appli
23	89.5	3.8	801	4	US-09-134-001C-5584	Sequence 5584, Ap
24	89.5	3.8	968	3	US-08-651-999A-7	Sequence 7, Appli
25	89.5	3.8	968	4	US-09-385-752-7	Sequence 7, Appli
26	89	3.8	1582	2	US-08-404-531B-9	Sequence 9, Appli
27	89	3.8	1582	3	US-08-476-900A-9	Sequence 9, Appli
28	89	3.8	1582	3	US-08-488-546A-9	Sequence 9, Appli
29	89	3.8	1582	4	US-08-726-320-5	Sequence 5, Appli
30	89	3.8	1582	4	US-09-208-716-5	Sequence 5, Appli
31	88.5	3.8	274	4	US-09-134-001C-3924	Sequence 3924, Ap
32	88.5	3.8	591	1	US-08-484-840-2	Sequence 2, Appli
33	88.5	3.8	591	1	US-08-483-094-2	Sequence 2, Appli
34	87	3.7	467	2	US-08-805-118-3	Sequence 3, Appli
35	87	3.7	467	4	US-09-391-958-3	Sequence 3, Appli
36	86.5	3.7	699	1	US-08-348-006B-7	Sequence 7, Appli
37	86.5	3.7	699	2	US-08-800-825A-7	Sequence 7, Appli
38	86.5	3.7	699	4	US-09-158-657-7	Sequence 7, Appli
39	85.5	3.6	861	3	US-09-022-875-2	Sequence 2, Appli
40	85	3.6	682	4	US-09-134-001C-5250	Sequence 5250, Ap
41	85	3.6	1580	4	US-08-726-320-1	Sequence 1, Appli
42	85	3.6	1580	4	US-09-208-716-1	Sequence 1, Appli
43	85	3.6	1581	2	US-08-404-531B-6	Sequence 6, Appli
44	85	3.6	1581	3	US-08-476-900A-6	Sequence 6, Appli
45	85	3.6	1581	3	US-08-488-546A-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-09-323-873A-8
; Sequence 8, Application US/09323873A
; Patent No. 6329503
; GENERAL INFORMATION:
; APPLICANT: Daniel E. Afar
; APPLICANT: Rene S. Hubert
; APPLICANT: Kahan Leong
; APPLICANT: Arthur B. Raitano
; APPLICANT: Douglas C. Saffran
; APPLICANT: Steve Chappell Mitchell
; TITLE OF INVENTION: NOVEL SERPENTINE TRANSMEMBRANE ANTIGENS
; TITLE OF INVENTION: EXPRESSED IN HUMAN CANCERS AND USES THEREOF
; FILE REFERENCE: 129.16USU2
; CURRENT APPLICATION NUMBER: US/09/323,873A
; CURRENT FILING DATE: 1999-06-01
; PRIOR APPLICATION NUMBER: 60/087,520
; PRIOR FILING DATE: 1998-06-01
; PRIOR APPLICATION NUMBER: 60/091,183
; PRIOR FILING DATE: 1998-06-30
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 173
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-09-323-873A-8

Query Match 38.3%; Score 901; DB 4; Length 173;
Best Local Similarity 100.0%; Pred. No. 5.8e-87;
Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    246 DFYKIPIEIVNKTLPIVAITLLSLVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQCRKQL 305
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 DFYKIPIEIVNKTLPIVAITLLSLVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQCRKQL 60

Qy    306 GLLSFFFAMVHVAYSLCLPMRRSERYLFLNMAYQQVHANIENSWNEEEVWRIEMYISFGI 365
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GLLSFFFAMVHVAYSLCLPMRRSERYLFLNMAYQQVHANIENSWNEEEVWRIEMYISFGI 120

Qy    366 MSLGLLSLLAVTSIPSVSNALNWREFSFIQSTLGYVALLISTFHVLIYGWKRA 418
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 MSLGLLSLLAVTSIPSVSNALNWREFSFIQSTLGYVALLISTFHVLIYGWKRA 173
```

RESULT 2
US-09-083-521-1
; Sequence 1, Application US/09083521
; Patent No. 6048970
; GENERAL INFORMATION:

```
;    APPLICANT:  Lal, Preeti
;    APPLICANT:  Guegler, Karl J.
;    APPLICANT:  Corley, Neil C.
;    TITLE OF INVENTION:  PROSTATE GROWTH-ASSOCIATED MEMBRANE PROTEINS
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  INCYTE PHARMACEUTICALS, INC.
;      STREET:  3174 PORTER DRIVE
;      CITY:  PALO ALTO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Word Perfect 6.1 for Windows/MS-DOS 6.2
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/083,521
;      FILING DATE:  Herewith
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CERRONE, MICHAEL C.
;      REGISTRATION NUMBER:  39,132
;      REFERENCE/DOCKET NUMBER:  PF-0527 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (650) 855-0555
;      TELEFAX:  (650) 845-4166
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  141 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  PROSTUT10
;      CLONE:  1691243
US-09-083-521-1

  Query Match             31.3%;  Score 736;  DB 3;  Length 141;
  Best Local Similarity   100.0%;  Pred. No. 1e-69;
  Matches  141;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     314 MVHVAYSLCLPMRRSERYLFLNMAYQQVHANIENSWNEEEVWRIEMYISFGIMSLGLLSL 373
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MVHVAYSLCLPMRRSERYLFLNMAYQQVHANIENSWNEEEVWRIEMYISFGIMSLGLLSL 60

Qy     374 LAVTSIPSVSNALNWREFSFIQSTLGYVALLISTFHVLIYGWKRAFEEEYYRFYTPPNFV 433
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 LAVTSIPSVSNALNWREFSFIQSTLGYVALLISTFHVLIYGWKRAFEEEYYRFYTPPNFV 120

Qy     434 LALVLPSIVILDLLQLCRYPD 454
          |||||||||||||||||||||
Db     121 LALVLPSIVILDLLQLCRYPD 141


RESULT 3
US-09-323-873A-2
; Sequence 2, Application US/09323873A
; Patent No. 6329503
; GENERAL INFORMATION:
;  APPLICANT: Daniel E. Afar
;  APPLICANT:  Rene S. Hubert
;  APPLICANT:  Kahan Leong
;  APPLICANT:  Arthur B. Raitano
;  APPLICANT:  Douglas C. Saffran
;  APPLICANT:  Steve Chappell Mitchell
;  TITLE OF INVENTION: NOVEL SERPENTINE TRANSMEMBRANE ANTIGENS
;  TITLE OF INVENTION:  EXPRESSED IN HUMAN CANCERS AND USES THEREOF
;  FILE REFERENCE: 129.16USU2
;  CURRENT APPLICATION NUMBER: US/09/323,873A
;  CURRENT FILING DATE:  1999-06-01
;  PRIOR APPLICATION NUMBER: 60/087,520
;  PRIOR FILING DATE: 1998-06-01
;  PRIOR APPLICATION NUMBER: 60/091,183
;  PRIOR FILING DATE: 1998-06-30
;  NUMBER OF SEQ ID NOS: 32
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 339
;   TYPE: PRT
;   ORGANISM: Homo Sapiens
US-09-323-873A-2

  Query Match             30.5%;  Score 717;  DB 4;  Length 339;
  Best Local Similarity   54.9%;  Pred. No. 3.9e-67;
  Matches  130;  Conservative   48;  Mismatches   59;  Indels    0;  Gaps    0;

Qy     208 LFTLWRGPVVVAISLATFFFLYSFVRDVIHPYARNQQSDFYKIPIEIVNKTLPIVAITLL 267
          ||  |  |: :|  :|:  |||: :|:|||| | : |  |||||| ::|| ||:|:||||
Db      67 LFPQWHLPIKIAAIIASLTFLYTLLREVIHPLATSHQQYFYKIPILVINKVLPMVSITLL 126

Qy     268 SLVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQCRKQLGLLSFFFAMVHVAYSLCLPMRR 327
          :|||| |::||  ||: ||||::|| ||: |:  ||| ||||||||::|  |||  ||||
Db     127 ALVYLPGVIAAIVQLHNGTKYKKFPHWLDKWMLTRKQFGLLSFFFAVLHAIYSLSYPMRR 186

Qy     328 SERYLFLNMAYQQVHANIENSWNEEEVWRIEMYISFGIMSLGLLSLLAVTSIPSVSNALN 387
          | ||  || |||||  | |::| | :|||:|:|:| ||: | :|:|||||||||||::|
Db     187 SYRYKLLNWAYQQVQQNKEDAWIEHDVWRMEIYVSLGIVGLAILALLAVTSIPSVSDSLT 246

Qy     388 WREFSFIQSTLGYVALLISTFHVLIYGWKRAFEEEYYRFYTPPNFVLALVLPSIVIL 444
          |||| :||| || |:||: | | ||: | :  : : : :|||| |::|: || :|::
Db     247 WREFHYIQSKLGIVSLLLGTIHALIFAWNKWIDIKQFVWYTPPTFMIAVFLPIVVLI 303


RESULT 4
US-09-655-270A-15
; Sequence 15, Application US/09655270A
; Patent No. 6329151
; GENERAL INFORMATION:
;  APPLICANT:  Rouviere, Pierre E.
;  TITLE OF INVENTION:  High Density Sampling of Differentially Expressed Prokaryotic
;  FILE REFERENCE:  BC1011 US NA
;  CURRENT APPLICATION NUMBER: US/09/655,270A
;  CURRENT FILING DATE:  2000-09-05
;  PRIOR APPLICATION NUMBER:  60/120,702
;  PRIOR FILING DATE:  1999-February-19
;  PRIOR APPLICATION NUMBER:  60/152,542
;  PRIOR FILING DATE:  1999-September-03
;  NUMBER OF SEQ ID NOS:  37
;  SOFTWARE:  Microsoft Office 97
; SEQ ID NO 15
;   LENGTH: 227
;   TYPE: PRT
;   ORGANISM: Rhodococcus erythropolis HL PM-1
US-09-655-270A-15

  Query Match              6.1%;  Score 144.5;  DB 4;  Length 227;
  Best Local Similarity   28.0%;  Pred. No. 4e-07;
  Matches   60;  Conservative   33;  Mismatches   88;  Indels   33;  Gaps    9;

Qy      25 IKDARKVTVGVIGSGDFAKSLTIRLIRCGYHVVIGSRNPKFASEFFPH---------VVD 75
          :| ::   ||  |:|   | |  |    |: ||||||: : | |             ||
Db       1 MKSSKIAVVG--GTGPQGKGLAYRFAAAGWPVVIGSRSAERAEEAALEVRRRAGDGAVVS 58

Qy      76 VTHHEDALTKTNIIFVAI----HREHYTSLWDLRHLLVGKILI--------DVSNNMRIN 123
            :  |     || : :    |||  :   :|  :  ||:::        | |     ::
Db      59 AADNASAAADCPIILLVVPYDGHRELVS---ELAPIFAGKLVVSCVNPLGFDKSGAYGLD 115

Qy     124 QYPESNAEYLASLFPDSLIVKGFNVVSA---WALQLGPKDASRQVYICSNNIQARQQVIE 180
            | || |  | | : :|  |: :||   |  : ||       | :|  ::  |: :|
Db     116 VEEGSAAEQLRDLVPGATVVAAFHHLSAVNLWEHE-GP--LPEDVLVCGDDRSAKDEVAR 172

Qy     181 LARQLNFIP-IDLGSLSSAREIENLPLRLFTLWR 213
          ||  :   | || |:|  ||::| |   |  : |
Db     173 LAVAITGRPGIDGGALRVARQLEPLTAVLINVNR 206
```

```
RESULT 5
US-09-651-941-17
; Sequence 17, Application US/09651941
; Patent No. 6355470
; GENERAL INFORMATION:
;  APPLICANT:  ROUVIER, PIERRE E
;  APPLICANT:  WALTERS, DANA M
;  APPLICANT:  RAINER, RUSS
;  TITLE OF INVENTION:  Genes Encoding Picric Acid Degradation
;  FILE REFERENCE:  BC1022 US NA
;  CURRENT APPLICATION NUMBER: US/09/651,941
;  CURRENT FILING DATE:  2000-08-31
;  PRIOR APPLICATION NUMBER:  60/152,545
;  PRIOR FILING DATE:  1999-10-03
;  NUMBER OF SEQ ID NOS:  28
;  SOFTWARE:  Microsoft Office 97
; SEQ ID NO 17
;   LENGTH: 227
;   TYPE: PRT
;   ORGANISM: Rhodococcus erythropolis HL PM-1
US-09-651-941-17

  Query Match             6.1%;  Score 144.5;  DB 4;  Length 227;
  Best Local Similarity   28.0%;  Pred. No. 4e-07;
  Matches   60;  Conservative   33;  Mismatches   88;  Indels   33;  Gaps    9;

Qy      25 IKDARKVTVGVIGSGDFAKSLTIRLIRCGYHVVIGSRNPKFASEFFPH---------VVD 75
           :| ::   ||  |:|   | |  |    |: ||||||: : | |             ||
Db       1 MKSSKIAVVG--GTGPQGKGLAYRFAAAGWPVVIGSRSAERAEEAALEVRRRAGDGAVVS 58

Qy      76 VTHHEDALTKTNIIFVAI----HREHYTSLWDLRHLLVGKILI--------DVSNNMRIN 123
             :  |     || : :    |||  :   :|  :  ||:::        | |    ::
Db      59 AADNASAAADCPIILLVVPYDGHRELVS---ELAPIFAGKLVVSCVNPLGFDKSGAYGLD 115

Qy     124 QYPESNAEYLASLFPDSLIVKGFNVVSA---WALQLGPKDASRQVYICSNNIQARQQVIE 180
           | || |  | | : :|  |: :||   |  : ||      | :| ::  |: :|
Db     116 VEEGSAAEQLRDLVPGATVVAAFHHLSAVNLWEHE-GP--LPEDVLVCGDDRSAKDEVAR 172

Qy     181 LARQLNFIP-IDLGSLSSAREIENLPLRLFTLWR 213
           ||  :   | || |:|  ||::| |   |  : |
Db     173 LAVAITGRPGIDGGALRVARQLEPLTAVLINVNR 206


RESULT 6
US-09-955-597-17
; Sequence 17, Application US/09955597
; Patent No. 6461856
; GENERAL INFORMATION:
;  APPLICANT: ROUVIER, PIERRE E
;  APPLICANT:  WALTERS, DANA M
;  APPLICANT:  RAINER, RUSS
;  TITLE OF INVENTION: Genes Encoding Picric Acid Degradation
;  FILE REFERENCE: BC1022 US NA
;  CURRENT APPLICATION NUMBER: US/09/955,597
;  CURRENT FILING DATE:  2001-09-17
;  PRIOR APPLICATION NUMBER: 60/152,545
;  PRIOR FILING DATE: 1999-10-03
;  NUMBER OF SEQ ID NOS: 28
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 17
;   LENGTH: 227
;   TYPE: PRT
;   ORGANISM: Rhodococcus erythropolis HL PM-1
US-09-955-597-17

  Query Match             6.1%;  Score 144.5;  DB 4;  Length 227;
  Best Local Similarity   28.0%;  Pred. No. 4e-07;
  Matches   60;  Conservative   33;  Mismatches   88;  Indels   33;  Gaps    9;

Qy      25 IKDARKVTVGVIGSGDFAKSLTIRLIRCGYHVVIGSRNPKFASEFFPH---------VVD 75
           :| ::   ||  |:|   | |  |    |: ||||||: : | |             ||
Db       1 MKSSKIAVVG--GTGPQGKGLAYRFAAAGWPVVIGSRSAERAEEAALEVRRRAGDGAVVS 58

Qy      76 VTHHEDALTKTNIIFVAI----HREHYTSLWDLRHLLVGKILI--------DVSNNMRIN 123
             :  |     || : :    |||  :   :|  :  ||:::        | |    ::
Db      59 AADNASAAADCPIILLVVPYDGHRELVS---ELAPIFAGKLVVSCVNPLGFDKSGAYGLD 115

Qy     124 QYPESNAEYLASLFPDSLIVKGFNVVSA---WALQLGPKDASRQVYICSNNIQARQQVIE 180
           | || |  | | : :|  |: :||   |  : ||      | :| ::  |: :|
Db     116 VEEGSAAEQLRDLVPGATVVAAFHHLSAVNLWEHE-GP--LPEDVLVCGDDRSAKDEVAR 172

Qy     181 LARQLNFIP-IDLGSLSSAREIENLPLRLFTLWR 213
           ||  :   | || |:|  ||::| |   |  : |
Db     173 LAVAITGRPGIDGGALRVARQLEPLTAVLINVNR 206


RESULT 7
US-08-487-886-2
; Sequence 2, Application US/08487886
; Patent No. 5744448
;  GENERAL INFORMATION:
;    APPLICANT:  Kelton, Christie Ann
;    APPLICANT:  Schweickhardt, Rene Lynn
;    APPLICANT:  Cheng, Shirley Vui Yen
;    APPLICANT:  Nugent, No. 5744448een Patrice
;    TITLE OF INVENTION:  Human Follicle Stimulating
;    TITLE OF INVENTION:  Hormone Receptor
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Stephan P. Williams,
;      ADDRESSEE:  Ares-Serono, Inc.
;      STREET:  Exchange Place, 37th floor
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" diskette, 1.44 MB, high density
;      COMPUTER:  IBM PS/2, model 55 SX
;      OPERATING SYSTEM:  MS-DOS version 4.0
;      SOFTWARE:  VAX/VMS Mass11 via Kermit to IBM MS-DOS
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/487,886
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/670,085
;      FILING DATE:  15-MAR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Williams, Stephan P.
;      REGISTRATION NUMBER:  28546
;      REFERENCE/DOCKET NUMBER:  US/252
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 723-1300
;      TELEFAX:  (617) 723-8923
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  695
;      TYPE:  Amino acid
;      TOPOLOGY:  Linear
;    MOLECULE TYPE:  protein
;    FEATURE:
;      NAME/KEY:  signal sequence
;      LOCATION:  -17 to -1
;      IDENTIFICATION METHOD:  hydrophobic
;    FEATURE:
;      NAME/KEY:  putative amino-terminal extracellular domain
;      LOCATION:  1 to 349
;      IDENTIFICATION METHOD:  similarity with other
;      IDENTIFICATION METHOD:  dimeric glycoprotein receptor extracellular
;      IDENTIFICATION METHOD:  domains, hydrophilic
;    FEATURE:
;      NAME/KEY:  transmembrane domain
;      LOCATION:  350 to 613
```

```
;       IDENTIFICATION METHOD:  similarity to other G
;       IDENTIFICATION METHOD:  protein-coupled receptor transmembrane domains
;     FEATURE:
;       NAME/KEY:  putative transmembrane region I
;       LOCATION:  350 to 370
;       IDENTIFICATION METHOD:  similarity to other G
;       IDENTIFICATION METHOD:  protein-coupled receptor transmembrane regions,
;       IDENTIFICATION METHOD:  hydrophobic, about 20-23 amino acids in length
;     FEATURE:
;       NAME/KEY:  putative transmembrane region II
;       LOCATION:  382 to 404
;       IDENTIFICATION METHOD:  similarity to other G
;       IDENTIFICATION METHOD:  protein-coupled receptor transmembrane regions,
;       IDENTIFICATION METHOD:  hydrophobic, about 20-23 amino acids in length
;     FEATURE:
;       NAME/KEY:  putative transmembrane region III
;       LOCATION:  427 to 448
;       IDENTIFICATION METHOD:  similarity to other G
;       IDENTIFICATION METHOD:  protein-coupled receptor transmembrane regions,
;       IDENTIFICATION METHOD:  hydrophobic, about 20-23 amino acids in length
;     FEATURE:
;       NAME/KEY:  putative transmembrane region IV
;       LOCATION:  469 to 491
;       IDENTIFICATION METHOD:  similarity to other G
;       IDENTIFICATION METHOD:  protein-coupled receptor transmembrane regions,
;       IDENTIFICATION METHOD:  hydrophobic, about 20-23 amino acids in length
;     FEATURE:
;       NAME/KEY:  putative transmembrane region V
;       LOCATION:  512 to 533
;       IDENTIFICATION METHOD:  similarity to other G
;       IDENTIFICATION METHOD:  protein-coupled receptor transmembrane regions,
;       IDENTIFICATION METHOD:  hydrophobic, about 20-23 amino acids in length
;     FEATURE:
;       NAME/KEY:  putative transmembrane region VI
;       LOCATION:  557 to 580
;       IDENTIFICATION METHOD:  similarity to other G
;       IDENTIFICATION METHOD:  protein-coupled receptor transmembrane regions,
;       IDENTIFICATION METHOD:  hydrophobic, about 20-23 amino acids in length
;     FEATURE:
;       NAME/KEY:  putative transmembrane region VII
;       LOCATION:  592 to 613
;       IDENTIFICATION METHOD:  similarity to other G
;       IDENTIFICATION METHOD:  protein-coupled receptor transmembrane regions,
;       IDENTIFICATION METHOD:  hydrophobic, about 20-23 amino acids in length
;     FEATURE:
;       NAME/KEY:  putative carboxy-terminal intracellular
;       NAME/KEY:  domain
;       LOCATION:  614 to 678
US-08-487-886-2

  Query Match             4.6%;  Score 107.5;  DB 1;  Length 695;
  Best Local Similarity   17.0%;  Pred. No. 0.018;
  Matches   84;  Conservative   77;  Mismatches  148;  Indels  185;  Gaps   17;

Qy      79 HEDALTKTNIIFVAIHREHYTSLWDLRHLLVG----------------KILIDVSNNMR 121
           ||  : | | : : |: | : :| :|::||:                |:|:|: :|:
Db      98 HEIRIEKANNL-LYINPEAFQNLPNLQYLLISNTGIKHLPDVHKIHSLQKVLLDIQDNIN 156

Qy     122 INQYPESNAEYLASLFPDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIEL 181
           |:           ::  :| :   |  |  |      |:  :::: |:
Db     157 IH----------TIERNSFVGLSFESVILWL----NKNGIQEIHNCA---------- 189

Qy     182 ARQLNFIPIDLGSLSSAREIENLPLRLFTLWRGPVVVAISLATFFFLYSFVRDVIHPYAR 241
             |   :|  :||    :| ||  :|    |||:: ||      | |:  : :
Db     190 ---FNGTQLDELNLSDNNNLEELPNDVFHGASGPVILDISRTRIHSLPSYGLENLKKLRA 246

Qy     242 NQQSDFYKIPIEIVNKTL-PIVAITLLSLVY---------------------------- 271
           :  |:|      ||  :||:   || |
Db     247 RSTYNLKKLP------TLEKLVALMEASLTYPSHCCAFANWRRQISELHPICNKSILRQE 300

Qy     272 ------------LAGLLAAAYQLYYGTKYRRF-------------PPWLETWLQCRKQL 305
                       ||    ::|   :   |  |             |  : :  |   :
Db     301 VDYMTQTRGQRSSLAEDNESSYSRGFDMTYTEFDYDLCNEVVDVTCSPKPDAFNPCEDIM 360
```

```
Qy     306 G-----LLSFFFAMVHVA------------YSLCLPMRRSERYLFLNMAYQQVHANIEN 347
           |     :| :| ::: :              | | :|     |:|  |:|:
Db     361 GYNILRVLIWFISILAITGNIIVLVILTTSQYKLTVP-----RFLMCNLAFAD------- 408

Qy     348 SWNEEEVWRIEMYISFGIMSLGLLSLLAVTSIPSVSNALNWRE---------FSFIQSTL 398
                     :  ||  | : |:   |      : |::|:          |:  | |
Db     409 ------------LCIGIYLLLIASVDIHTKSQYHNYAIDWQTGAGCDAAGFFTVFASEL 455

Qy     399 G---YVALLISTFH------------------VLIYGWKRAFEEEYYRFYTPPNFV-LA 435
             |: :  :|                    |:: ||  ||    :  :   ::: ::
Db     456 SVYTLTAITLERWHTITHAMQLDCKVQLRHAASVMVMGWIFAFAAALFPIFGISSYMKVS 515

Qy     436 LVLPSIVILDLLQL 449
           : ||  :   | ||
Db     516 ICLPMDIDSPLSQL 529
```

```
RESULT 8
US-08-482-855-2
; Sequence 2, Application US/08482855
; Patent No. 6121016
;  GENERAL INFORMATION:
;    APPLICANT:  Kelton, Christie Ann
;    APPLICANT:  Schweickhardt, Rene Lynn
;    APPLICANT:  Cheng, Shirley Vui Yen
;    APPLICANT:  Nugent, No. 6121016een Patrice
;    TITLE OF INVENTION:  Human Follicle Stimulating
;    TITLE OF INVENTION:  Hormone Receptor
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Stephan P. Williams,
;      ADDRESSEE:  Ares-Serono, Inc.
;      STREET:  Exchange Place, 37th floor
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" diskette, 1.44 MB, high density
;      COMPUTER:  IBM PS/2, model 55 SX
;      OPERATING SYSTEM:  MS-DOS version 4.0
;      SOFTWARE:  VAX/VMS Mass11 via Kermit to IBM MS-DOS
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/482,855
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/670,085
;      FILING DATE:  15-MAR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Williams, Stephan P.
;      REGISTRATION NUMBER:  28546
;      REFERENCE/DOCKET NUMBER:  US/252
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 723-1300
;      TELEFAX:  (617) 723-8923
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  695
;      TYPE:  Amino acid
;      TOPOLOGY:  Linear
;    MOLECULE TYPE:  protein
;    FEATURE:
;      NAME/KEY:  signal sequence
;      LOCATION:  -17 to -1
;      IDENTIFICATION METHOD:  hydrophobic
;    FEATURE:
;      NAME/KEY:  putative amino-terminal extracellular domain
;      LOCATION:  1 to 349
;      IDENTIFICATION METHOD:  similarity with other
;      IDENTIFICATION METHOD:  dimeric glycoprotein receptor extracellular
;      IDENTIFICATION METHOD:  domains, hydrophilic
```

```
;     FEATURE:
;       NAME/KEY:  transmembrane domain
;       LOCATION:  350 to 613
;       IDENTIFICATION METHOD:  similarity to other G
;       IDENTIFICATION METHOD:  protein-coupled receptor transmembrane domains
;     FEATURE:
;       NAME/KEY:  putative transmembrane region I
;       LOCATION:  350 to 370
;       IDENTIFICATION METHOD:  similarity to other G
;       IDENTIFICATION METHOD:  protein-coupled receptor transmembrane regions,
;       IDENTIFICATION METHOD:  hydrophobic, about 20-23 amino acids in length
;     FEATURE:
;       NAME/KEY:  putative transmembrane region II
;       LOCATION:  382 to 404
;       IDENTIFICATION METHOD:  similarity to other G
;       IDENTIFICATION METHOD:  protein-coupled receptor transmembrane regions,
;       IDENTIFICATION METHOD:  hydrophobic, about 20-23 amino acids in length
;     FEATURE:
;       NAME/KEY:  putative transmembrane region III
;       LOCATION:  427 to 448
;       IDENTIFICATION METHOD:  similarity to other G
;       IDENTIFICATION METHOD:  protein-coupled receptor transmembrane regions,
;       IDENTIFICATION METHOD:  hydrophobic, about 20-23 amino acids in length
;     FEATURE:
;       NAME/KEY:  putative transmembrane region IV
;       LOCATION:  469 to 491
;       IDENTIFICATION METHOD:  similarity to other G
;       IDENTIFICATION METHOD:  protein-coupled receptor transmembrane regions,
;       IDENTIFICATION METHOD:  hydrophobic, about 20-23 amino acids in length
;     FEATURE:
;       NAME/KEY:  putative transmembrane region V
;       LOCATION:  512 to 533
;       IDENTIFICATION METHOD:  similarity to other G
;       IDENTIFICATION METHOD:  protein-coupled receptor transmembrane regions,
;       IDENTIFICATION METHOD:  hydrophobic, about 20-23 amino acids in length
;     FEATURE:
;       NAME/KEY:  putative transmembrane region VI
;       LOCATION:  557 to 580
;       IDENTIFICATION METHOD:  similarity to other G
;       IDENTIFICATION METHOD:  protein-coupled receptor transmembrane regions,
;       IDENTIFICATION METHOD:  hydrophobic, about 20-23 amino acids in length
;     FEATURE:
;       NAME/KEY:  putative transmembrane region VII
;       LOCATION:  592 to 613
;       IDENTIFICATION METHOD:  similarity to other G
;       IDENTIFICATION METHOD:  protein-coupled receptor transmembrane regions,
;       IDENTIFICATION METHOD:  hydrophobic, about 20-23 amino acids in length
;     FEATURE:
;       NAME/KEY:  putative carboxy-terminal intracellular
;       NAME/KEY:  domain
;       LOCATION:  614 to 678
US-08-482-855-2

  Query Match             4.6%;  Score 107.5;  DB 3;  Length 695;
  Best Local Similarity   17.0%;  Pred. No. 0.018;
  Matches   84;  Conservative   77;  Mismatches  148;  Indels  185;  Gaps   17;

Qy      79 HEDALTKTNIIFVAIHREHYTSLWDLRHLLVG----------------KILIDVSNNMR 121
           ||  : | | : : |: | : :| :|::||:                 |:|:|: :|:
Db      98 HEIRIEKANNL-LYINPEAFQNLPNLQYLLISNTGIKHLPDVHKIHSLQKVLLDIQDNIN 156

Qy     122 INQYPESNAEYLASLFPDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIEL 181
           |:          ::  :| :   |  |  |      |:  :::: |:
Db     157 IH----------TIERNSFVGLSFESVILWL----NKNGIQEIHNCA---------- 189

Qy     182 ARQLNFIPIDLGSLSSAREIENLPLRLFTLWRGPVVVAISLATFFFLYSFVRDVIHPYAR 241
               |  :|  :||    :| ||  :|    |||:: ||      | |:  : :
Db     190 ---FNGTQLDELNLSDNNNLEELPNDVFHGASGPVILDISRTRIHSLPSYGLENLKKLRA 246

Qy     242 NQQSDFYKIPIEIVNKTL-PIVAITLLSLVY---------------------------- 271
           :  |:|      ||  :||:   || |
Db     247 RSTYNLKKLP------TLEKLVALMEASLTYPSHCCAFANWRRQISELHPICNKSILRQE 300

Qy     272 ------------LAGLLAAAYQLYYGTKYRRF------------PPWLETWLQCRKQL 305
                       ||    ::|   :   | |              |  : :  |   :
Db     301 VDYMTQTRGQRSSLAEDNESSYSRGFDMTYTEFDYDLCNEVVDVTCSPKPDAFNPCEDIM 360

Qy     306 G-----LLSFFFAMVHVA------------YSLCLPMRRSERYLFLNMAYQQVHANIEN 347
           |     :| :| ::: :             | | :|     |:|  |:|:
Db     361 GYNILRVLIWFISILAITGNIIVLVILTTSQYKLTVP-----RFLMCNLAFAD------- 408

Qy     348 SWNEEEVWRIEMYISFGIMSLGLLSLLAVTSIPSVSNALNWRE---------FSFIQSTL 398
                     :  ||  | : |:   |     : |::|:          |:   | |
Db     409 ------------LCIGIYLLLIASVDIHTKSQYHNYAIDWQTGAGCDAAGFFTVFASEL 455

Qy     399 G---YVALLISTFH------------------VLIYGWKRAFEEEYYRFYTPPNFV-LA 435
             |: :  :|                   |:: ||  ||    :  :   ::: ::
Db     456 SVYTLTAITLERWHTITHAMQLDCKVQLRHAASVMVMGWIFAFAAALFPIFGISSYMKVS 515

Qy     436 LVLPSIVILDLLQL 449
           : ||  :   | ||
Db     516 ICLPMDIDSPLSQL 529


RESULT 9
US-08-474-986-2
;   GENERAL INFORMATION:
;       APPLICANT: Kelton, Christie Ann
;                  Schweickhardt, Rene Lynn
;                  Cheng, Shirley Vui Yen
;                  Nugent, No. 6372711een Patrice
;       TITLE OF INVENTION: Human Follicle Stimulating
;                           Hormone Receptor
;       NUMBER OF SEQUENCES: 2
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Stephan P. Williams,
;                       Ares-Serono, Inc.
;            STREET: Exchange Place, 37th floor
;            CITY: Boston
;            STATE: MA
;            COUNTRY: USA
;            ZIP: 02109
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: 3.5" diskette, 1.44 MB, high density
;            COMPUTER: IBM PS/2, model 55 SX
;            OPERATING SYSTEM: MS-DOS version 4.0
;            SOFTWARE: VAX/VMS Mass11 via Kermit to IBM MS-DOS
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/08/474,986
;            FILING DATE: 07-Jun-1995
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 07/670,085
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Williams, Stephan P.
;            REGISTRATION NUMBER: 28546
;            REFERENCE/DOCKET NUMBER: US/252
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (617) 723-1300
;            TELEFAX: (617) 723-8923
;            LOCATION:  614 to 678
;       SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-08-474-986-2

  Query Match             4.6%;  Score 107.5;  DB 4;  Length 695;
  Best Local Similarity   17.0%;  Pred. No. 0.018;
  Matches   84;  Conservative   77;  Mismatches  148;  Indels  185;  Gaps   17;

Qy      79 HEDALTKTNIIFVAIHREHYTSLWDLRHLLVG----------------KILIDVSNNMR 121
           ||  : | | : : |: | : :| :|::||:                 |:|:|: :|:
Db      98 HEIRIEKANNL-LYINPEAFQNLPNLQYLLISNTGIKHLPDVHKIHSLQKVLLDIQDNIN 156

Qy     122 INQYPESNAEYLASLFPDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIEL 181
           |:          ::  :| :   |  |  |      |:  :::: |:
Db     157 IH----------TIERNSFVGLSFESVILWL----NKNGIQEIHNCA---------- 189
```

```
Qy    182 ARQLNFIPIDLGSLSSAREIENLPLRLFTLWRGPVVVAISLATFFFLYSFVRDVIHPYAR 241
Db    190 ---FNGTQLDELNLSDNNNLEELPNDVFHGASGPVILDISRTRIHSLPSYGLENLKKLRA 246

Qy    242 NQQSDFYKIPIEIVNKTL-PIVAITLLSLVY---------------------------- 271
Db    247 RSTYNLKKLP------TLEKLVALMEASLTYPSHCCAFANWRRQISELHPICNKSILRQE 300

Qy    272 ------------LAGLLAAAYQLYYGTKYRRF------------PPWLETWLQCRKQL 305
Db    301 VDYMTQTRGQRSSLAEDNESSYSRGFDMTYTEFDYDLCNEVVDVTCSPKPDAFNPCEDIM 360

Qy    306 G-----LLSFFFAMVHVA------------YSLCLPMRRSERYLFLNMAYQQVHANIEN 347
Db    361 GYNILRVLIWFISILAITGNIIVLVILTTSQYKLTVP-----RFLMCNLAFAD------- 408

Qy    348 SWNEEEVWRIEMYISFGIMSLGLLSLLAVTSIPSVSNALNWRE---------FSFIQSTL 398
Db    409 ------------LCIGIYLLLIASVDIHTKSQYHNYAIDWQTGAGCDAAGFFTVFASEL 455

Qy    399 G---YVALLISTFH------------------VLIYGWKRAFEEEYYRFYTPPNFV-LA 435
Db    456 SVYTLTAITLERWHTITHAMQLDCKVQLRHAASVMVMGWIFAFAAALFPIFGISSYMKVS 515

Qy    436 LVLPSIVILDLLQL 449
Db    516 ICLPMDIDSPLSQL 529
```

RESULT 10
US-09-323-873A-20
; Sequence 20, Application US/09323873A
; Patent No. 6329503
; GENERAL INFORMATION:
; APPLICANT: Daniel E. Afar
; APPLICANT: Rene S. Hubert
; APPLICANT: Kahan Leong
; APPLICANT: Arthur B. Raitano
; APPLICANT: Douglas C. Saffran
; APPLICANT: Steve Chappell Mitchell
; TITLE OF INVENTION: NOVEL SERPENTINE TRANSMEMBRANE ANTIGENS
; TITLE OF INVENTION: EXPRESSED IN HUMAN CANCERS AND USES THEREOF
; FILE REFERENCE: 129.16USU2
; CURRENT APPLICATION NUMBER: US/09/323,873A
; CURRENT FILING DATE: 1999-06-01
; PRIOR APPLICATION NUMBER: 60/087,520
; PRIOR FILING DATE: 1998-06-01
; PRIOR APPLICATION NUMBER: 60/091,183
; PRIOR FILING DATE: 1998-06-30
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 34
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: STEAP-1 PEPTIDE
US-09-323-873A-20

Query Match 4.6%; Score 107; DB 4; Length 34;
Best Local Similarity 58.8%; Pred. No. 0.00019;
Matches 20; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

```
Qy    326 RRSERYLFLNMAYQQVHANIENSWNEEEVWRIEM 359
Db      1 RRSYRYKLLNWAYQQVQQNKEDAWIEHDVWRMEI 34
```

RESULT 11
US-09-316-083-3
; Sequence 3, Application US/09316083A
; Patent No. 6280942
; GENERAL INFORMATION:
; APPLICANT: The Institute of Physical and Chemical Research
; TITLE OF INVENTION: Endonuclease
; FILE REFERENCE: PH-651
; CURRENT APPLICATION NUMBER: US/09/316,083A
; CURRENT FILING DATE: 1999-05-20
; EARLIER APPLICATION NUMBER: JP98/141861
; EARLIER FILING DATE: 1998-05-22
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-316-083-3

Query Match 4.3%; Score 101; DB 4; Length 476;
Best Local Similarity 20.7%; Pred. No. 0.047;
Matches 52; Conservative 48; Mismatches 85; Indels 66; Gaps 12;

```
Qy     62 NPKFASEFFPHVVDVTHHEDALTKTNI----IFVAIHREHYT----------SLWD 103
Db    254 NPYFVNAFSINI----------KTNLAKEKIFTNIYNKLYSDYKINQINNHIPYYNYLK 302

Qy    104 LRHLLVGKILIDVSNNMRINQYPESNAEYLASLF--PDSLIVKGFNVVSAWALQLGPKDA 161
Db    303 INNKLPIKNIMDIKNNYWLAGFTAADGSFLSSMYNPKDTLLFKNM---------- 347

Qy    162 SRQVYICSNNIQARQQVIELARQLNFIPIDLGSLSSAREIENLPLRLFTLWRGPVVVAIS 221
Db    348 -RPSYVIS-QVETRKELIYLIQE----SFDL-SISNVKKVGNRKLKDFKLFTRTTDELMK 400

Qy    222 LATFFFLYSFVRDVIHPYARNQQSDFYKIP----IEIVNKTLPIVAITLLSLVYLAGLLA 277
Db    401 -----FIYYF--DKFLPLHDNKQFNYIKFRFNTFIKSYNWNNRVFGLVLSE--YINNIKI 451

Qy    278 AAYQLYYGTKY 288
Db    452 DNYDYYYYNKY 462
```

RESULT 12
US-08-724-974A-2
; Sequence 2, Application US/08724974A
; Patent No. 5912335
; GENERAL INFORMATION:
; APPLICANT: Derk J. Bergsma, Catherine E. Ellis
; TITLE OF INVENTION: A No. 5912335el G-Protein Coupled Receptor
; TITLE OF INVENTION: HUVCT36
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road, P.O. Box 1539
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
; COMPUTER: IBM 486
; OPERATING SYSTEM: WINDOWS FOR WORKGROUPS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,974A
; FILING DATE: October 3, 1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: William T. Han
; REGISTRATION NUMBER: 34,344
; REFERENCE/DOCKET NUMBER: ATG50022

```
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610 270 5024
;      TELEFAX:  610 270 5090
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  531
;      TYPE:  Amino Acid
;      TOPOLOGY:  Linear
US-08-724-974A-2

  Query Match             4.2%;  Score 99;  DB 2;  Length 531;
  Best Local Similarity   22.6%;  Pred. No. 0.091;
  Matches   48;  Conservative   37;  Mismatches   75;  Indels   52;  Gaps    9;

Qy     255 VNKTL-PIVAITLLSLVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQCRKQLGLLSFFFA 313
           :::|| |:| :|:|    : |  |    ||:|        :|:  :: | :||:
Db     132 IHQTLAPVVYVTVL----VVGFPANCLSLYFG--------YLQ--IKARNELGVYLCNLT 177

Qy     314 MVHVAYSLCLPMRRSERYLFLNMAYQQVHANIENSWNEEEVWRIEMYISFGIMSLGLLSL 373
           :  : |   ||           : |   | |  :     :|  | :| :  | |:| |
Db     178 VADLFYICSLPFW---------LQYVLQHDNWSHGDLSCQVCGILLYENIYI-SVGFLCC 227

Qy     374 LAVTSIPSVSNALNWREFSFIQSTLGYVA------LLISTF------------------- 408
           ::|    :|::   : :|  ::: :|         || | :
Db     228 ISVDRYLAVAHPFRFHQFRTLKAAVGVSVVIWAKELLTSIYFLMHEEVIEDENQHRVCFE 287

Qy     409 HVLIYGWKRAFEEEYYRFYTPPNFVLALVLPS 440
           |  |  |:||    ||||     | : |:| |
Db     288 HYPIQAWQRAI--NYYRFLVGFLFPICLLLAS 317


RESULT 13
US-08-460-576-2
; Sequence 2, Application US/08460576
; Patent No. 6033903
;  GENERAL INFORMATION:
;    APPLICANT:  SISK, William P.
;    TITLE OF INVENTION:  A METHOD OF EXPRESSING AND SECRETING
;    TITLE OF INVENTION:  SOLUBLE EXTRACELLULAR DOMAINS OF HUMAN GONADOTROPIN
;    TITLE OF INVENTION:  HORMONE RECEPTORS
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BROWDY AND NEIMARK
;      STREET:  419 Seventh Street, N.W., Suite 300
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/460,576
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  YUN, Allen C.
;      REGISTRATION NUMBER:  37,971
;      REFERENCE/DOCKET NUMBER:  SISK=1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-628-5197
;      TELEFAX:  202-737-3528
;      TELEX:  248633
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  390 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-460-576-2

  Query Match             4.1%;  Score 97.5;  DB 3;  Length 390;
  Best Local Similarity   21.3%;  Pred. No. 0.081;
  Matches   45;  Conservative   39;  Mismatches   72;  Indels   55;  Gaps    7;

Qy      79 HEDALTKTNIIFVAIHREHYTSLWDLRHLLVG----------------KILIDVSNNMR 121
           ||  : | | : : |: | : :| :|::||:                 |:|:|: :|:
Db     122 HEIRIEKANNL-LYINPEAFQNLPNLQYLLISNTGIKHLPDVHKIHSLQKVLLDIQDNIN 180

Qy     122 INQYPESNAEYLASLFPDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIEL 181
           |:           ::  :| :   |  |  |      |:  :::: |:
Db     181 IH----------TIERNSFVGLSFESVILWL----NKNGIQEIHNCA----------- 213

Qy     182 ARQLNFIPIDLGSLSSAREIENLPLRLFTLWRGPVVVAISLATFFFLYSFVRDVIHPYAR 241
               |  :|  :||    :| ||  :|    |||:: ||      | |:  : :
Db     214 ---FNGTQLDELNLSDNNNLEELPNDVFHGASGPVILDISRTRIHSLPSYGLENLKKLRA 270

Qy     242 NQQSDFYKIPIEIVNKTL-PIVAITLLSLVY 271
           :  |:|      ||  :||:   || |
Db     271 RSTYNLKKLP------TLEKLVALMEASLTY 295


RESULT 14
US-07-757-342D-6
; Sequence 6, Application US/07757342D
; Patent No. 6218509
;   GENERAL INFORMATION:
;       APPLICANT: IGARASHI, Masao
;                  MINEGISHI, Takashi
;                  NAKAMURA, Kazuto
;       TITLE OF INVENTION: PROTEIN, DNA AND USE THEREOF
;       NUMBER OF SEQUENCES: 10
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
;                       CUSHMAN
;            STREET: 130 Water Street
;            CITY: Boston
;            STATE: Massachusetts
;            COUNTRY: US
;            ZIP: 02109
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.25
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/07/757,342D
;            FILING DATE: 10-Sep-1991
;            CLASSIFICATION: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: BUCKLEY, Linda M.
;            REGISTRATION NUMBER: 31003
;            REFERENCE/DOCKET NUMBER: 41226
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (617)523-3400
;            TELEFAX: (617)523-6440
;            TELEX: 200291 STRE UR
;   INFORMATION FOR SEQ ID NO: 6:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 692 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-07-757-342D-6

  Query Match             4.1%;  Score 97;  DB 4;  Length 692;
  Best Local Similarity   17.3%;  Pred. No. 0.22;
  Matches   86;  Conservative   70;  Mismatches  148;  Indels  194;  Gaps   20;

Qy      79 HEDALTKTNIIFVAIHREHYTSLWDLRHLLVG----------------KILIDVSNNMR 121
           ||  : | | : : |: | : :|  ||:||:                 |:|:|: :|:
Db      98 HEIRIEKANNL-LYINPEAFQNLPSLRYLLISNTGIKHLPAVHKIQSLQKVLLDIQDNIN 156
```

```
Qy     122 INQYPESNAEYLASLFPDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIEL 181
           |:           :  :| :   |  |  |      |:   ::: |:
Db     157 IH----------IVARNSFMGLSFESVILWL----SKNGIEEIHNCA---------- 189

Qy     182 ARQLNFIPIDLGSLSSAREIENLPLRLFTLWRGPVVVAISLATFFFLYSFVRDVIHPYAR 241
               |   :|  :||    :| ||  :|    |||:: ||          |  :|
Db     190 ---FNGTQLDELNLSDNNNLEELPNDVFQGASGPVILDIS----------RTKVHSLPN 235

Qy     242 NQQSDFYKIPIEIVN--KTLP----IVAITLLSLVY------LAGLLAAAYQLY------ 283
           :   :  |:       | ||     | :   || |       | |     :|:
Db     236 HGLENLKKLRARSTYRLKKLPNLDKFVTLMEASLTYPSHCCAFANLKRQISELHPICNKS 295

Qy     284 ------------------------YG----TKYRRF------------PPWLETWLQC 301
                                    ||      |  |            |  : :  |
Db     296 ILRQDIDDMTQIGDQRVSLIDDEPSYGKGSDMMYNEFDYDLCNEVVDVTCSPKPDAFNPC 355

Qy     302 RKQLG-----LLSFFFAMVHVA------------YSLCLPMRRSERYLFLNMAYQQVHA 343
             :|     :| :| ::: :              | | :|     |:|  |:|:
Db     356 EDIMGYNILRVLIWFISILAITGNTTVLVVLTTSQYKLTVP-----RFLMCNLAFAD--- 407

Qy     344 NIENSWNEEEVWRIEMYISFGIMSLGLLSLLAVTSIPSVSNALNWRE---------FSFI 394
                       :  ||  | : |:   |      : |::|:          |:
Db     408 ----------------LCIGIYLLLIASVDIHTKSQYHNYAIDWQTGAGCDAAGFFTVF 450

Qy     395 QSTLG---YVALLISTFH------------------VLIYGWKRAFEEEYYRFYTPPNF 432
           | |      |: :  :|                   |:: ||  ||   :  :   ::
Db     451 ASELSVYTLTAITLERWHTITHAMQLECKVQLRHAASVMVLGWTFAFAAALFPIFGISSY 510

Qy     433 V-LALVLPSIVILDLLQL 449
           : ::: ||  :   | ||
Db     511 MKVSICLPMDIDSPLSQL 528
```

RESULT 15
US-08-724-974A-3
; Sequence 3, Application US/08724974A
; Patent No. 5912335
; GENERAL INFORMATION:
; APPLICANT: Derk J. Bergsma, Catherine E. Ellis
; TITLE OF INVENTION: A No. 5912335el G-Protein Coupled Receptor
; TITLE OF INVENTION: HUVCT36
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road, P.O. Box 1539
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
; COMPUTER: IBM 486
; OPERATING SYSTEM: WINDOWS FOR WORKGROUPS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,974A
; FILING DATE: October 3, 1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: William T. Han
; REGISTRATION NUMBER: 34,344
; REFERENCE/DOCKET NUMBER: ATG50022
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610 270 5024
; TELEFAX: 610 270 5090
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 365
; TYPE: Amino Acid
; TOPOLOGY: Linear

US-08-724-974A-3

Query Match 4.0%; Score 93; DB 2; Length 365;
Best Local Similarity 22.2%; Pred. No. 0.22;
Matches 47; Conservative 37; Mismatches 76; Indels 52; Gaps 9;

```
Qy     255 VNKTL-PIVAITLLSLVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQCRKQLGLLSFFFA 313
           :::|| |:| :|:|     : |  |    ||:|        :|:  :: | :||:
Db      19 IHQTLAPVVYVTVL----VVGFPANCLSLYFG--------YLQ--IKARNELGVYLCNLT 64

Qy     314 MVHVAYSLCLPMRRSERYLFLNMAYQQVHANIENSWNEEEVWRIEMYISFGIMSLGLLSL 373
           :  : |   ||           : |   | |  :     :|  | :| :  | |:| |
Db      65 VADLFYICSLPFW---------LQYVLQHDNWSHGDLSCQVCGILLYENIYI-SVGFLCC 114

Qy     374 LAVTSIPSVSNALNWREFSFIQSTLGYVA------LLISTF------------------ 408
           ::|    :|::   : :|  ::: :           || | :
Db     115 ISVDRYLAVAHPFRFHQFRTLKAAVRVTVVIWAKELLTSIYFLMHEEVIEDENQHRVCFE 174

Qy     409 HVLIYGWKRAFEEEYYRFYTPPNFVLALVLPS 440
           |  |  |:||    ||||      | : |:| |
Db     175 HYPIQAWQRAI--NYYRFLVGFLFPICLLLAS 204
```

Search completed: January 17, 2003, 18:59:36
Job time : 23 secs

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: January 17, 2003, 18:58:14 ; Search time 36 Seconds
(without alignments)
2598.483 Million cell updates/sec

Title: US-09-455-486-6
Perfect score: 2351
Sequence: 1 MESISMMGSPKSLSETCLPN.....ALVLPSIVILDLLQLCRYPD 454

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1316	56.0	488	11	Q99P41	Q99p41 rattus norv
2	1306	55.6	514	11	Q924Z1	Q924z1 mus musculu
3	1269	54.0	488	4	Q9NVB5	Q9nvb5 homo sapien
4	1245.5	53.0	487	4	Q8TF03	Q8tf03 homo sapien
5	1074.5	45.7	456	4	Q8TDP3	Q8tdp3 homo sapien
6	1064.5	45.3	470	11	Q923B6	Q923b6 mus musculu
7	1057.5	45.0	470	11	Q91W31	Q91w31 mus musculu
8	1041.5	44.3	474	11	Q91ZE8	Q91ze8 mus musculu
9	728	31.0	338	6	Q9GL50	Q9gl50 sus scrofa
10	715	30.4	339	11	Q924Z2	Q924z2 mus musculu
11	707	30.1	339	11	Q9CWR7	Q9cwr7 mus musculu
12	704	29.9	339	11	Q924J9	Q924j9 mus musculu
13	628	26.7	283	4	Q9H5R1	Q9h5r1 homo sapien
14	577	24.5	264	4	Q9Y6U5	Q9y6u5 homo sapien
15	541	23.0	283	4	Q8WWB0	Q8wwb0 homo sapien
16	426	18.1	132	11	Q9D5R1	Q9d5r1 mus musculu
17	324	13.8	143	4	Q9H7Y1	Q9h7y1 homo sapien
18	170	7.2	208	16	Q8XQS2	Q8xqs2 ralstonia s
19	160	6.8	198	16	Q930K7	Q930k7 rhizobium m
20	159.5	6.8	211	16	Q8YK44	Q8yk44 anabaena sp
21	156.5	6.7	239	16	Q9RKR6	Q9rkr6 streptomyce
22	151	6.4	200	16	Q98C62	Q98c62 rhizobium l
23	147	6.3	213	17	O29059	O29059 archaeoglob
24	144.5	6.1	226	2	Q9AH05	Q9ah05 rhodococcus
25	143.5	6.1	223	17	Q8TYQ9	Q8tyq9 methanopyru
26	142	6.0	212	17	O29370	O29370 archaeoglob
27	142	6.0	224	1	O59661	O59661 methanobact
28	137.5	5.8	199	2	Q9F418	Q9f418 mycobacteri
29	136.5	5.8	217	16	Q9JN78	Q9jn78 streptomyce
30	128.5	5.5	220	2	Q8VPU7	Q8vpu7 nocardioide
31	121	5.1	191	16	Q9CF67	Q9cf67 lactococcus
32	119.5	5.1	216	2	O50637	O50637 leptospira
33	119.5	5.1	254	16	Q8UKE0	Q8uke0 agrobacteri
34	117.5	5.0	429	17	Q970N3	Q970n3 sulfolobus
35	116.5	5.0	242	16	Q9PCP1	Q9pcp1 xylella fas
36	114	4.8	207	16	Q9RIU0	Q9riu0 streptomyce
37	112.5	4.8	207	16	Q9RIT8	Q9rit8 streptomyce
38	112.5	4.8	247	17	Q8TIC0	Q8tic0 methanosarc
39	112	4.8	408	5	Q9V8W8	Q9v8w8 drosophila
40	111.5	4.7	1228	3	Q12754	Q12754 saccharomyc
41	110.5	4.7	222	17	Q9HMC5	Q9hmc5 halobacteri
42	110	4.7	341	17	Q96Y98	Q96y98 sulfolobus
43	109	4.6	219	16	Q9RIU2	Q9riu2 streptomyce
44	109	4.6	320	12	Q9YVM4	Q9yvm4 melanoplus
45	109	4.6	1301	10	O23530	O23530 arabidopsis

ALIGNMENTS

RESULT 1
Q99P41
ID Q99P41 PRELIMINARY; PRT; 488 AA.
AC Q99P41;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Tumor suppressor pHyde.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DUNNING; TISSUE=PROSTATE CANCER TUMOR;
RX MEDLINE=20424188; PubMed=10969787;
RA Steiner M.S., Zhang X., Wang Y., Lu Y.;
RT "Growth inhibition of prostate cancer by an adenovirus expressing a
RT novel tumor suppressor gene, pHyde.";
RL Cancer Res. 60:4419-4425(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=DUNNING; TISSUE=PROSTATE CANCER TUMOR;
RA Lu Y., Rinaldy A.R., Steiner M.S.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Steiner M.S., Rinaldy A.;
RT "A novel prostate cancer associated gene.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF335281; AAK00361.1; -.
DR EMBL; AF238865; AAL78207.1; -.
SQ SEQUENCE 488 AA; 54640 MW; 59FF07121919FDCB CRC64;

Query Match 56.0%; Score 1316; DB 11; Length 488;
Best Local Similarity 56.2%; Pred. No. 9.9e-97;
Matches 248; Conservative 81; Mismatches 100; Indels 12; Gaps 4;

Qy 6 MMGSPKSLSETCLPNGINGIKDARKVTVGVIGSGDFAKSLTIRLIRCGYHVVIGSRNPKF 65

```
          ::  |  ||:|  :|      |:| |  ||::||||||:||  ||:  |: ||:||||||
Db     14 LVDSDGSLAE--VP------KEAPK--VGILGSGDFARSLATRLVGSGFFVVVGSRNPKR 63

Qy     66 ASEFFPHVVDVTHHEDALTKTNIIFVAIHREHYTSLWDLRHLLVGKILIDVSNNMRIN-- 123
          :  || :  ||  |:|::   :||||: ||||:||  |   | ||||:||||
Db     64 TAGLFPSLAQVTFQEEAVSSPEVIFVAVFREHYSSLCSLADQLAGKILVDVSNPTEKERL 123

Qy    124 QYPESNAEYLASLFPDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIELAR 183
          |: :|||||||||||   :|| |||:|||||| ||:| :||| || : ::|:  | |:||
Db    124 QHRQSNAEYLASLFPACTVVKAFNVISAWALQAGPRDGNRQVLICGDQLEAKHTVSEMAR 183

Qy    184 QLNFIPIDLGSLSSAREIENLPLRLFTLWRGPVVVAISLATFFFLYSFVRDVIHPYARNQ 243
          : | |:|:|||:||||:| :|||||   |: | ::|: |:|  : |:|:|||: || |
Db    184 AMGFTPLDMGSLASAREVEAIPLRLLPSWKVPTLLALGLSTQSYAYNFIRDVLQPYIRKD 243

Qy    244 QSDFYKIPIEIVNKTLPIVAITLLSLVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQCRK 303
          :: |||:|: :|| |:| ||  |||||||  |:|||| ||  ||||:||| ||: ||| ||
Db    244 ENKFYKMPLSVVNTTIPCVAYVLLSLVYLPGVLAAALQLRRGTKYQRFPDWLDHWLQHRK 303

Qy    304 QLGLLSFFFAMVHVAYSLCLPMRRSERYLFLNMAYQQVHANIENSWNEEEVWRIEMYISF 363
          |:|||||||||:|  || |||:||| ||  :|:| :|| ||    | ||||||:|:|:|
Db    304 QIGLLSFFFAMLHALYSFCLPLRRSHRYDLVNLAVKQVLANKSRLWVEEEVWRMEIYLSL 363

Qy    364 GIMSLGLLSLLAVTSIPSVSNALNWREFSFIQSTLGYVALLISTFHVLIYGWKRAFEEEY 423
          |:::||:|||||||||||::|:|||:||||:|||||:|||::|| | | ||| |||||  :
Db    364 GVLALGMLSLLAVTSIPSIANSLNWKEFSFVQSTLGFVALMLSTMHTLTYGWTRAFEENH 423

Qy    424 YRFYTPPNFVLALVLPSIVIL 444
          |:|| || | | |:|| ::||
Db    424 YKFYLPPTFTLTLLLPCVIIL 444


RESULT 2
Q924Z1
ID   Q924Z1      PRELIMINARY;      PRT;   514 AA.
AC   Q924Z1;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Dudulin 2.
GN   1010001D01RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Serru V., Lamblin D., Lenoir C., Manivet P., Vaubourdolle M.,
RA   Kellermann O., Loric S.;
RT   "Molecular cloning and expression of mouse dudulin 2.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY029586; AAK50539.1; -.
DR   MGD; MGI:1915678; 1010001D01Rik.
SQ   SEQUENCE   514 AA;  57268 MW;  339886C288AEC0E2 CRC64;

  Query Match           55.6%;  Score 1306;  DB 11;  Length 514;
  Best Local Similarity   56.2%;  Pred. No. 6.6e-96;
  Matches  248;  Conservative   81;  Mismatches  100;  Indels   12;  Gaps    4;

Qy      6 MMGSPKSLSETCLPNGINGIKDARKVTVGVIGSGDFAKSLTIRLIRCGYHVVIGSRNPKF 65
          ::  |  ||:|  :|      |:| |  ||::||||||:||  ||:  |: ||:||||||
Db     14 LVDSDGSLAE--VP------KEAPK--VGILGSGDFARSLATRLVGSGFSVVVGSRNPKR 63

Qy     66 ASEFFPHVVDVTHHEDALTKTNIIFVAIHREHYTSLWDLRHLLVGKILIDVSNNMRIN-- 123
          :  || :  ||  |:|::   :||||: ||||:||  |   | ||||:||||
Db     64 TAGLFPSLAQVTFQEEAVSSPEVIFVAVFREHYSSLCSLADQLAGKILVDVSNPTEKEHL 123

Qy    124 QYPESNAEYLASLFPDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIELAR 183
          |: :|||||||||||   :|| |||:|||||| ||:| :||| || |: :|:|:|:|:||
Db    124 QHRQSNAEYLASLFPACTVVKAFNVISAWALQAVPRDGNRQVLICGNDSKAKQRVMEMAR 183

Qy    184 QLNFIPIDLGSLSSAREIENLPLRLFTLWRGPVVVAISLATFFFLYSFVRDVIHPYARNQ 243
          : | |:|:|||:||||:| :|||||   : | ::|: |   |: |:|:|||: || |
```

```
Db    184 AMGFTPLDMGSLASAREVEAIPLRLLPSGKVPTLLALGLFVCFYTYNFIRDVLQPYIRKD 243

Qy    244 QSDFYKIPIEIVNKTLPIVAITLLSLVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQCRK 303
          :: |||:|: :|| |:| ||  |||||||  |:|||| ||  ||||:||| ||: ||| ||
Db    244 ENKFYKMPLSVVNTTLPCVAYVLLSLVYLPGVLAAALQLRRGTKYQRFPDWLDHWLQHRK 303

Qy    304 QLGLLSFFFAMVHVAYSLCLPMRRSERYLFLNMAYQQVHANIENSWNEEEVWRIEMYISF 363
          |:|||||||||:|  || |||:||| ||  :|:| :|| ||    | ||||||:|:|:|
Db    304 QIGLLSFFFAMLHALYSFCLPLRRSHRYDLVNLAVKQVLANKSRLWVEEEVWRMEIYLSL 363

Qy    364 GIMSLGLLSLLAVTSIPSVSNALNWREFSFIQSTLGYVALLISTFHVLIYGWKRAFEEEY 423
          |:::||:|||||||||||::|:|||:||||:|||||:|||::|| | | ||| |||||  :
Db    364 GVLALGMLSLLAVTSLPSIANSLNWKEFSFVQSTLGFVALILSTMHTLTYGWTRAFEENH 423

Qy    424 YRFYTPPNFVLALVLPSIVIL 444
          |:|| || | | |:|| ::||
Db    424 YKFYLPPTFTLTLLLPCVIIL 444


RESULT 3
Q9NVB5
ID   Q9NVB5      PRELIMINARY;      PRT;   488 AA.
AC   Q9NVB5;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CDNA FLJ10829 fis, clone NT2RP4001138 (Dudulin 2).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA   Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA   Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA   Nakamura Y., Nagahari K., Masuho Y., Sasaki N.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Serru V., Manivet P., Lenoir C., Eschwege P., Lamblin D.,
RA   Vaubourdolle M., Kellermann O., Loric S.;
RT   "Dudulin 2, a new tumor antigen expressed in various human tumors.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK001691; BAA91839.1; -.
DR   EMBL; AY029585; AAK50538.1; -.
DR   InterPro; IPR003006; Ig_MHC.
DR   PROSITE; PS00290; IG_MHC; UNKNOWN_1.
SQ   SEQUENCE   488 AA;  54616 MW;  BC0BCA483335AAD6 CRC64;

  Query Match           54.0%;  Score 1269;  DB 4;  Length 488;
  Best Local Similarity   54.3%;  Pred. No. 5.5e-93;
  Matches  241;  Conservative   83;  Mismatches  108;  Indels   12;  Gaps    4;

Qy      3 SISMMGSPKSLSETCLPNGINGIKDARKVTVGVIGSGDFAKSLTIRLIRCGYHVVIGSRN 62
          |: :: |  ||::  :|:      :| |  ||::||||||:||  ||:  |: ||:||||
Db     11 SLHLVDSDSSLAK--VPD------EAPK--VGILGSGDFARSLATRLVGSGFKVVVGSRN 60

Qy     63 PKFASEFFPHVVDVTHHEDALTKTNIIFVAIHREHYTSLWDLRHLLVGKILIDVSN--NM 120
          ||  :  :|   ||  :|::   :||||: ||||:||  |   | ||||:||||  ||
Db     61 PKRTARLYPSAAQVTFQEEAVSSPEVIFVAVFREHYSSLCSLSDQLAGKILVDVSNPTEQ 120

Qy    121 RINQYPESNAEYLASLFPDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIE 180
            |: |||||||||||   :|| |||:||| || ||:| :||| || :  :|:: | |
Db    121 EHLQHRESNAEYLASLFPTCTVVKAFNVISAWTLQAGPRDGNRQVPICGDQPEAKRAVSE 180

Qy    181 LARQLNFIPIDLGSLSSAREIENLPLRLFTLWRGPVVVAISLATFFFLYSFVRDVIHPYA 240
          :|  : |:|:|:|||:|| |:| :||||   |: | ::|: |   |: |:|||||: ||
Db    181 MALAMGFMPVDMGSLASAWEVEAMPLRLLPAWKVPTLLALGLFVCFYAYNFVRDVLQPYV 240

Qy    241 RNQQSDFYKIPIEIVNKTLPIVAITLLSLVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQ 300
          :  |: |:|:|: :|| ||| ||  ||||||| |:|||| ||  ||||:||| ||: |||
```

```
Db    241 QESQNKFFKLPVSVVNTTLPCVAYVLLSLVYLPGVLAAALQLRRGTKYQRFPDWLDHWLQ 300

Qy    301 CRKQLGLLSFFFAMVHVAYSLCLPMRRSERYLFLNMAYQQVHANIENSWNEEEVWRIEMY 360
          |||:|||||| | :|  || |||:||: ||  :|:| :|| ||  : | ||||||:|:|
Db    301 HRKQIGLLSFFCAALHALYSFCLPLRRAHRYDLVNLAVKQVLANKSHLWVEEEVWRMEIY 360

Qy    361 ISFGIMSLGLLSLLAVTSIPSVSNALNWREFSFIQSTLGYVALLISTFHVLIYGWKRAFE 420
          :| |:::|| |||||||:||:::|:|||||||:||:||:|||::|| | | ||| ||||
Db    361 LSLGVLALGTLSLLAVTSLPSIANSLNWREFSFVQSSLGFVALVLSTLHTLTYGWTRAFE 420

Qy    421 EEYYRFYTPPNFVLALVLPSIVIL 444
          |  |:|| || | | |::| :|||
Db    421 ESRYKFYLPPTFTLTLLVPCVVIL 444


RESULT 4
Q8TF03
ID   Q8TF03      PRELIMINARY;      PRT;   487 AA.
AC   Q8TF03;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Tumor suppressor pHyde.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Steiner M.S., Allay J.A., Wang C.;
RT   "A novel prostate-derived tumor suppressor.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF238864; AAL78206.1; -.
SQ   SEQUENCE   487 AA;  54402 MW;  AFF16053590E6F68 CRC64;

  Query Match             53.0%;  Score 1245.5;  DB 4;  Length 487;
  Best Local Similarity   53.8%;  Pred. No. 4.1e-91;
  Matches  239;  Conservative   83;  Mismatches  109;  Indels   13;  Gaps    5;

Qy      3 SISMMGSPKSLSETCLPNGINGIKDARKVTVGVIGSGDFAKSLTIRLIRCGYHVVIGSRN 62
          |: :: |  ||::  :|:      :| ||:  ::|||||:|| ||:  |: ||:||||
Db     11 SLHLVDSDSSLAK--VPD------EAPKVS--ILGSGDFARSLATRLVGSGFKVVVGSRN 60

Qy     63 PKFASEFFPHVVDVTHHEDALTKTNIIFVAIHREHYTSLWDLRHLLVGKILIDVSN--NM 120
          ||  :  ||    ||  |:|::   :||||: ||||:||  |   | ||||:||||
Db     61 PKRTARLFPSAAQVTFQEEAVSSPEVIFVAVFREHYSSLCSLSDQLAGKILVDVSNPTEQ 120

Qy    121 RINQYPESNAEYLASLFPDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIE 180
            |: ||||||||||||   :|| |||:||| || ||:| : || || :  :|:: | |
Db    121 EHLQHRESNAEYLASLFPTCTVVKAFNVISAWTLQAGPRDGNGQVPICGDQPEAKRAVSE 180

Qy    181 LARQLNFIPIDLGSLSSAREIENLPLRLFTLWRGPVVVAISLATFFFLYSFVRDVIHPYA 240
          :|  : |:|:|:|||:|| |:| :||||   |:.| ::|: |   |: |:|||||: ||
Db    181 MALAMGFMPVDMGSLASAWEVEAMPLRLLPAWKVPTLLALGLFVCFYAYNFVRDVLQPYV 240

Qy    241 RNQQSDFYKIPIEIVNKTLPIVAITLLSLVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQ 300
          :  |: |:|:|: :|| ||| ||  ||||||| |:|||| ||  ||||:||| ||: |||
Db    241 QESQNKFFKLPVSVVNTTLPCVAYVLLSLVYLPGVLAAALQLRRGTKYQRFPDWLDHWLQ 300

Qy    301 CRKQLGLLSFFFAMVHVAYSLCLPMRRSERYLFLNMAYQQVHANIENSWNEEEVWRIEMY 360
          |||:|||||| | :|  || |||:||: ||  :|:| :|| ||  : | |||||:|:|
Db    301 HRKQIGLLSFFCAALHALYSFCLPLRRAHRYDLVNLAVKQVLANKSHLWVEEVWRMEIY 359

Qy    361 ISFGIMSLGLLSLLAVTSIPSVSNALNWREFSFIQSTLGYVALLISTFHVLIYGWKRAFE 420
          :| |:::|| |||||||:||:::|:|||||||:||:||:|||::|| | | ||| ||||
Db    360 LSLGVLALGTLSLLAVTSLPSIANSLNWREFSFVQSSLGFVALVLSTLHTLTYGWTRAFE 419

Qy    421 EEYYRFYTPPNFVLALVLPSIVIL 444
          |  |:|| || | | |::| :|||
Db    420 ESRYKFYLPPTFTLTLLVPCVVIL 443


RESULT 5
Q8TDP3
ID   Q8TDP3      PRELIMINARY;      PRT;   456 AA.
AC   Q8TDP3;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   PHYDE II.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wang C., Allay J.A., Steiner M.S.;
RT   "Second human member of pHyde family, Human pHyde II.";
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF262322; AAM08128.1; -.
SQ   SEQUENCE   456 AA;  50430 MW;  C5F7C7008D55251E CRC64;

  Query Match             45.7%;  Score 1074.5;  DB 4;  Length 456;
  Best Local Similarity   52.9%;  Pred. No. 1.6e-77;
  Matches  209;  Conservative   75;  Mismatches   98;  Indels   13;  Gaps    5;

Qy      3 SISMMGSPKSLSETCLPNGINGIKDARKVTVGVIGSGDFAKSLTIRLIRCGYHVVIGSRN 62
          |: :: |  ||::  :|:      :| ||:  ::|||||:|| ||:  |: ||:||||
Db     11 SLHLVDSDSSLAK--VPD------EAPKVS--ILGSGDFARSLATRLVGSGFKVVVGSRN 60

Qy     63 PKFASEFFPHVVDVTHHEDALTKTNIIFVAIHREHYTSLWDLRHLLVGKILIDVSN--NM 120
          ||  :  ||    ||  |:|::   :||||: ||||:||  |   | ||||:||||
Db     61 PKRTARLFPSAAQVTFQEEAVSSPEVIFVAVFREHYSSLCSLSDQLAGKILVDVSNPTEQ 120

Qy    121 RINQYPESNAEYLASLFPDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIE 180
            |: ||||||||||||   :|| |||:||| || ||:| : || || :  :|:: | |
Db    121 EHLQHRESNAEYLASLFPTCTVVKAFNVISAWTLQAGPRDGNGQVPICGDQPEAKRAVSE 180

Qy    181 LARQLNFIPIDLGSLSSAREIENLPLRLFTLWRGPVVVAISLATFFFLYSFVRDVIHPYA 240
          :|  : |:|:|:|||:|| |:| :||||   |:.| ::|: |   |: |:|||||: ||
Db    181 MALAMGFMPVDMGSLASAWEVEAMPLRLLPAWKVPTLLALGLFVCFYAYNFVRDVLQPYV 240

Qy    241 RNQQSDFYKIPIEIVNKTLPIVAITLLSLVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQ 300
          :  |: |:|:|: :|| ||| ||  ||||||| |:|||| ||  ||||:||| ||: |||
Db    241 QESQNKFFKLPVSVVNTTLPCVAYVLLSLVYLPGVLAAALQLRRGTKYQRFPDWLDHWLQ 300

Qy    301 CRKQLGLLSFFFAMVHVAYSLCLPMRRSERYLFLNMAYQQVHANIENSWNEEEVWRIEMY 360
          |||:|||||| | :|  || |||:||: ||  :|:| :|| ||  : | |||||:|:|
Db    301 HRKQIGLLSFFCAALHALYSFCLPLRRAHRYDLVNLAVKQVLANKSHLW-VEEVWRMEIY 359

Qy    361 ISFGIMSLGLLSLLAVTSIPSVSNALNWREFSFIQ 395
          :| |:::|| |||||||:||::|:|||||||:|
Db    360 LSLGVLALGTLSLLAVTSLPSIANSLNWREFSFVQ 394


RESULT 6
Q923B6
ID   Q923B6      PRELIMINARY;      PRT;   470 AA.
AC   Q923B6;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Similar to hypothetical protein FLJ23153.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC006651; AAH06651.1; -.
DR   InterPro; IPR000585; Hemopexin.
DR   PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
SQ   SEQUENCE   470 AA;  52994 MW;  6823E7682AC78B19 CRC64;
```

```
  Query Match           45.3%;  Score 1064.5;  DB 11;  Length 470;
  Best Local Similarity  48.5%;  Pred. No. 1.1e-76;
  Matches  200;  Conservative   80;  Mismatches  131;  Indels    1;  Gaps    1;

Qy      33 VGVIGSGDFAKSLTIRLIRCGYHVVIGSRNPKFASEFFPHVVDVTHHEDALTKTNIIFVA 92
           | : |:||| ||| :::::||| :| |||||:  |   |   :|  : :| :|::|| :|
Db      22 VCIFGTGDFGKSLGLKMLQCGYSIVFGSRNPQ-VSSLLPRGAEVLSYSEAASKSDIIILA 80

Qy      93 IHREHYTSLWDLRHLLVGKILIDVSNNMRINQYPESNAEYLASLFPDSLIVKGFNVVSAW 152
           :|||||  || :|   | ||:|:|||||| :|||||||||||||| | | : :|| || :|||
Db      81 MHREHYDSLTELVDYLKGKVLVDVSNNRKINQYPESNAEYLAQLEPGAHVVKAFNTISAW 140

Qy     153 ALQLGPKDASRQVYICSNNIQARQQVIELARQLNFIPIDLGSLSSAREIENLPLRLFTLW 212
           ||| |  |||||||::| |: :|:|:|:::|| |   |:| ||| :| |||| ||:|| :|
Db     141 ALQSGTLDASRQVFVCGNDSKAKQRVMDIARTLGLTPLDQGSLMAASEIENYPLQLFPMW 200

Qy     213 RGPVVVAISLATFFFLYSFVRDVIHPYARNQQSDFYKIPIEIVNKTLPIVAITLLSLVYL 272
           | |  ::  |  |||:|  :|:||:||   :    |:: | | |:  || |: ||:||||
Db     201 RFPFYLSSVLCVFFFVYCAIREVIYPYVNGKTDATYRLAISIPNRVFPITALILLALVYL 260

Qy     273 AGLLAAAYQLYYGTKYRRFPPWLETWLQCRKQLGLLSFFFAMVHVAYSLCLPMRRSERYL 332
           |:|||  ||| |||||||| ||: |: |||||||::  || :|| |:| :|:|   |:
Db     261 PGILAAILQLYRGTKYRRFPNWLDHWMLCRKQLGLVALGFAFLHVIYTLVIPIRYYVRWR 320

Qy     333 FLNMAYQQVHANIENSWNEEEVWRIEMYISFGIMSLGLLSLLAVTSIPSVSNALNWREFS 392
           |    |   | :: :    |  : |:: ||:   |  || :||:|||| :|||||
Db     321 LRNATITQALTNKDSPFITSYAWINDSYLALGILGFFLFLLLGITSLPSVSNMVNWREFR 380

Qy     393 FIQSTLGYVALLISTFHVLIYGWKRAFEEEYYRFYTPPNFVLALVLPSIVIL 444
           |:|| |||: |:: | | |:|| ||       |:  |  ::||||:|  |::
Db     381 FVQSKLGYLTLVLCTAHTLVYGGKRFLSPSILRWSLPSAYILALVIPCAVLV 432
```

RESULT 7
Q91W31
ID Q91W31 PRELIMINARY; PRT; 470 AA.
AC Q91W31;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Transmembrane protein TIARP.
GN TIARP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SWISS; TISSUE=ADIPOSE TISSUE;
RX MEDLINE=21424005; PubMed=11443137;
RA Moldes M., Lasnier F., Gauthereau X., Klein C., Pairault J., Feve B.,
RA Chambaut-Guerin A.M.;
RT "Tumor necrosis factor-alpha-induced Adipose-related protein (TIARP),
RT a cell-surface protein that is highly induced by tumor necrosis
RT factor-alpha and adipose conversion.";
RL J. Biol. Chem. 276:33938-33946(2001).
DR EMBL; AJ319746; CAC41351.1; -.
DR MGD; MGI:2152335; Tiarp.
DR InterPro; IPR000585; Hemopexin.
DR PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
SQ SEQUENCE 470 AA; 52971 MW; F19239188E08663E CRC64;

```
  Query Match           45.0%;  Score 1057.5;  DB 11;  Length 470;
  Best Local Similarity  48.1%;  Pred. No. 3.8e-76;
  Matches  198;  Conservative   81;  Mismatches  132;  Indels    1;  Gaps    1;

Qy      33 VGVIGSGDFAKSLTIRLIRCGYHVVIGSRNPKFASEFFPHVVDVTHHEDALTKTNIIFVA 92
           | : |:||| ||| :::::||| :| |||||:  |   |   :|  : :| :|::|| :|
Db      22 VCIFGTGDFGKSLGLKMLQCGYSIVFGSRNPQ-VSSLLPRGAEVLSYSEAASKSDIIILA 80

Qy      93 IHREHYTSLWDLRHLLVGKILIDVSNNMRINQYPESNAEYLASLFPDSLIVKGFNVVSAW 152
           :|||||  || :|   | ||:|:|||||| :|||||||||||||| | | : :|| || :|||
Db      81 MHREHYDSLTELVDCLKGKVLVDVSNNRKINQYPESNAEYLAQLEPGAHVVKAFNTISAW 140

Qy     153 ALQLGPKDASRQVYICSNNIQARQQVIELARQLNFIPIDLGSLSSAREIENLPLRLFTLW 212
           ||| |  |||||||::| |: :|:|:|:::|| |   |:| ||| :| |||| ||:|| :|
Db     141 ALQSGTLDASRQVFVCGNDSKAKQRVMDIARTLGLTPLDQGSLMAASEIENYPLQLFPMW 200

Qy     213 RGPVVVAISLATFFFLYSFVRDVIHPYARNQQSDFYKIPIEIVNKTLPIVAITLLSLVYL 272
           | |  ::  |  |||:|  :|:||:||   :    |:: | | |:  || |: ||:||||
Db     201 RFPFYLSSVLCIFFFVYCAIREVIYPYVNGKTDATYRLAISIPNRVFPITALILLALVYL 260

Qy     273 AGLLAAAYQLYYGTKYRRFPPWLETWLQCRKQLGLLSFFFAMVHVAYSLCLPMRRSERYL 332
           |:|||  ||| |||||||| ||: |: |||||||::  || :|| |:| :|:|   |:
Db     261 PGILAAILQLYRGTKYRRFPNWLDHWMLCRKQLGLVALGFAFLHVIYTLVIPIRYYVRWR 320

Qy     333 FLNMAYQQVHANIENSWNEEEVWRIEMYISFGIMSLGLLSLLAVTSIPSVSNALNWREFS 392
           |    |   | :: :    |  : |:: ||:   |  || :||:|||| :|||||
Db     321 LRNATITQALTNKDSPFITSYAWINDSYLALGILGFFLFLLLGITSLPSVSNMVNWREFR 380

Qy     393 FIQSTLGYVALLISTFHVLIYGWKRAFEEEYYRFYTPPNFVLALVLPSIVIL 444
           |:|| |||: |:: | | |:|| ||       |:  |  ::||||:|  |::
Db     381 FVQSKLGYLTLVLCTAHTLVYGGKRFLSPSILRWSLPSAYILALVIPCAVLV 432
```

RESULT 8
Q91ZE8
ID Q91ZE8 PRELIMINARY; PRT; 474 AA.
AC Q91ZE8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Dudulin 4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Serru V., Lamblin D., Manivet P., Pernet P., Vaubourdolle M.,
RA Kellermann O., Loric S.;
RT "Molecular cloning and expression of two new members of the dudulin
RT family.";
RL Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AY029778; AAK40270.1; -.
DR InterPro; IPR000585; Hemopexin.
DR PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
SQ SEQUENCE 474 AA; 53543 MW; 937EEA3393D13956 CRC64;

```
  Query Match           44.3%;  Score 1041.5;  DB 11;  Length 474;
  Best Local Similarity  47.8%;  Pred. No. 7.3e-75;
  Matches  197;  Conservative   79;  Mismatches  135;  Indels    1;  Gaps    1;

Qy      33 VGVIGSGDFAKSLTIRLIRCGYHVVIGSRNPKFASEFFPHVVDVTHHEDALTKTNIIFVA 92
           | : |:||| ||| :::::||| :| |||||:  |   |   :|  : :| :|::|| :|
Db      22 VCIFGTGDFGKSLGMKMLQCGYSIVFGSRNPQ-VSSLLPRGAEVLSYSEAASKSDIIILA 80

Qy      93 IHREHYTSLWDLRHLLVGKILIDVSNNMRINQYPESNAEYLASLFPDSLIVKGFNVVSAW 152
           :|||||  || :|   | ||:|:|||||| :|||||||||||||| | | : :|| || :|||
Db      81 MHREHYDSLTELVDYLKGKVLVDVSNNHKINQYPESNAEYLAQLEPEAHVVKAFNTISAW 140

Qy     153 ALQLGPKDASRQVYICSNNIQARQQVIELARQLNFIPIDLGSLSSAREIENLPLRLFTLW 212
           ||| |  |||||| :| |: :|:|:|:::|| |   |:| ||| :| |||| ||:|| :|
Db     141 ALQSGTLDASRQVLVCGNDSKAKQRVMDIARTLGLTPLDQGSLMAASEIENYPLQLFPMW 200

Qy     213 RGPVVVAISLATFFFLYSFVRDVIHPYARNQQSDFYKIPIEIVNKTLPIVAITLLSLVYL 272
           | |  ::  |  |||||  :|:||:||   :    |:: | | |:  || |: ||:||||
Db     201 RFPFYLSSVLCIFFFLYCAIREVIYPYMNGKTDATYRLAISIPNRVFPITALILLALVYL 260

Qy     273 AGLLAAAYQLYYGTKYRRFPPWLETWLQCRKQLGLLSFFFAMVHVAYSLCLPMRRSERYL 332
           |:|||  ||| |||||||| ||: |: |||||||::  || :|  |:| :|:|   |:
Db     261 PGILAAILQLYRGTKYRRFPNWLDHWMLCRKQLGLVALGFAFLHAIYTLVIPIRYYVRWR 320

Qy     333 FLNMAYQQVHANIENSWNEEEVWRIEMYISFGIMSLGLLSLLAVTSIPSVSNALNWREFS 392
           |    |   | :: :    |  : |:: ||:   |  |: :||:|||| :|||||
Db     321 LRKATITQALTSKRQPFITSYAWINDSYLALGILGFFLFLLVGITSLPSVSNMVNWREFR 380
```

```
Qy     393 FIQSTLGYVALLISTFHVLIYGWKRAFEEEYYRFYTPPNFVLALVLPSIVIL 444
           |:|| |||: |:: | | |:|| ||       |:  |  ::||||:|  |::
Db     381 FVQSKLGYLTLVLCTAHTLVYGGKRFLSPSILRWSLPSAYILALVIPCAVLV 432


RESULT 9
Q9GL50
ID   Q9GL50      PRELIMINARY;      PRT;   338 AA.
AC   Q9GL50;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   Six transmembrane endothelial antigen of PAEC.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Nagasaka T., Boulday G., Coupel S., Coulon F., Tesson L.,
RA   Heslan J.-M., Soulillou J.-P., Charreau B.;
RT   "Differential gene expression in endothelial cells during TNF-alpha-
RT   and LPS-mediated activation.";
RL   Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF319659; AAG33868.1; -.
KW   Transmembrane.
SQ   SEQUENCE   338 AA;  39918 MW;  ED490E86E067A32B CRC64;

  Query Match             31.0%;  Score 728;  DB 6;  Length 338;
  Best Local Similarity   49.3%;  Pred. No. 4.7e-50;
  Matches  136;  Conservative   59;  Mismatches   73;  Indels    8;  Gaps    2;

Qy     170 NNIQARQQVIELARQLNFIPIDL-GSLSSAREIENLPLRLFTLWRGPVVVAISLATFFFL 228
           |::  |  :: | :  :|   |    |   :|       ||  |  |: :|  :::  ||
Db      34 NSMPKRPMLVHLHQTAHFDEFDCPPELQHKQE-------LFPKWHLPIKIAAIVSSLTFL 86

Qy     229 YSFVRDVIHPYARNQQSDFYKIPIEIVNKTLPIVAITLLSLVYLAGLLAAAYQLYYGTKY 288
           |: :|:||||:  : |  |||||| ::|| ||:|:||||:|||| |::||  ||: ||||
Db      87 YTLLREVIHPFVTSHQQYFYKIPILVINKVLPMVSITLLALVYLPGVIAAIVQLHNGTKY 146

Qy     289 RRFPPWLETWLQCRKQLGLLSFFFAMVHVAYSLCLPMRRSERYLFLNMAYQQVHANIENS 348
           ::|| ||: |:  ||| |||||||::|  |||  ||||| ||  || |||||  | |::
Db     147 KKFPHWLDRWMVTRKQFGLLSFFFAVLHAVYSLSYPMRRSYRYKLLNWAYQQVQQNKEDA 206

Qy     349 WNEEEVWRIEMYISFGIMSLGLLSLLAVTSIPSVSNALNWREFSFIQSTLGYVALLISTF 408
           | | :|||:|:|:| ||::| :|:|||||||||||::| |||| :|||||| |:|| |
Db     207 WIEHDVWRMEIYVSLGIVTLAILALLAVTSIPSVSDSLTWREFHYIQSTLGIVSLLLGTI 266

Qy     409 HVLIYGWKRAFEEEYYRFYTPPNFVLALVLPSIVIL 444
           | ||: | :  : : : :|||| |::|: ||::|::
Db     267 HALIFAWNKWVDIKQFIWYTPPTFMIAVFLPTVVLI 302


RESULT 10
Q924Z2
ID   Q924Z2      PRELIMINARY;      PRT;   339 AA.
AC   Q924Z2;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Dudulin.
GN   1010001D01RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Serru V., Manivet P., Lamblin D., Vaubourdolle M., Kellermann O.,
RA   Loric S.;
RT   "Prostate and non-prostate expression of dudulin, the mouse ortholog
RT   of human STEAP.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY029584; AAK50537.1; -.
DR   MGD; MGI:1915678; 1010001D01Rik.
SQ   SEQUENCE   339 AA;  39109 MW;  32A2C29F2E333BD0 CRC64;

  Query Match             30.4%;  Score 715;  DB 11;  Length 339;
  Best Local Similarity   54.5%;  Pred. No. 5.1e-49;
  Matches  132;  Conservative   49;  Mismatches   61;  Indels    0;  Gaps    0;

Qy     209 FTLWRGPVVVAISLATFFFLYSFVRDVIHPYARNQQSDFYKIPIEIVNKTLPIVAITLLS 268
           |  || || ||  :::  |||: :|::|:|   :::  |||||| :|| ||| ||  |||
Db      68 FPNWRLPVKVAAIISSLTFLYTLLREIIYPLVTSREQYFYKIPILVVNTTLPCVAYVLLS 127

Qy     269 LVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQCRKQLGLLSFFFAMVHVAYSLCLPMRRS 328
           |||| |:|||| ||  ||||:||| ||: ||| |||:|||||||||:|  || |||:|||
Db     128 LVYLPGVLAAALQLRRGTKYQRFPDWLDHWLQHRKQIGLLSFFFAMLHALYSFCLPLRRS 187

Qy     329 ERYLFLNMAYQQVHANIENSWNEEEVWRIEMYISFGIMSLGLLSLLAVTSIPSVSNALNW 388
           ||  :|:|: :|| || |::| | :|||:|:|:| ||: | :|:|||||||||||::| |
Db     188 HRYDLVNLAVKQVLANKEDAWVEHDVWRMEIYVSLGIVGLAILALLAVTSIPSVSDSLTW 247

Qy     389 REFSFIQSTLGYVALLISTFHVLIYGWKRAFEEEYYRFYTPPNFVLALVLPSIVILDLLQ 448
           ||| :||| || |:||: | | |:: | :  :   : :| || |::|: ||::|::  :
Db     248 REFHYIQSKLGIVSLLLGTVHALVFAWNKWVDVSQFVWYMPPTFMIAVFLPTLVLICKIA 307

Qy     449 LC 450
           ||
Db     308 LC 309


RESULT 11
Q9CWR7
ID   Q9CWR7      PRELIMINARY;      PRT;   339 AA.
AC   Q9CWR7;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   2410007B19Rik protein.
GN   STEAP OR 2410007B19RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=EMBRYONIC STEM CELLS;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK010437; BAB26938.1; -.
DR   MGD; MGI:1917608; Steap.
SQ   SEQUENCE   339 AA;  39264 MW;  3F7AB9C7520F0968 CRC64;

  Query Match             30.1%;  Score 707;  DB 11;  Length 339;
  Best Local Similarity   52.9%;  Pred. No. 2.2e-48;
```

```
  Matches  128;  Conservative   51;  Mismatches   63;  Indels    0;  Gaps    0;

Qy     209 FTLWRGPVVVAISLATFFFLYSFVRDVIHPYARNQQSDFYKIPIEIVNKTLPIVAITLLS 268
Db      68 FPNWRLPVKVAAIISSLTFLYTLLREIIYPLVTSREQYFYKIPILVINKVLPMVAITLLA 127

Qy     269 LVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQCRKQLGLLSFFFAMVHVAYSLCLPMRRS 328
Db     128 LVYLPGELAAVVQLRNGTKYKKFPPWLDRWMLAKKQFGLLSFFFAVLHAVYSLSYPMRRS 187

Qy     329 ERYLFLNMAYQQVHANIENSWNEEEVWRIEMYISFGIMSLGLLSLLAVTSIPSVSNALNW 388
Db     188 YRYKLLNWAYKQVQQNKEDAWVEHDVWRMEIYVSLGIVGLAILALLAVTSIPSVSDSLTW 247

Qy     389 REFSFIQSTLGYVALLISTFHVLIYGWKRAFEEEYYRFYTPPNFVLALVLPSIVILDLLQ 448
Db     248 REFHYIQSKLGIVSLLLGTVHALVFAWNKWVDVSQFVWYMPPTFMIAVFLPTLVLICKIA 307

Qy     449 LC 450
Db     308 LC 309


RESULT 12
Q924J9
ID   Q924J9      PRELIMINARY;      PRT;   339 AA.
AC   Q924J9;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Six-transmembrane epithelial antigen of the prostate.
GN   STEAP.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6;
RX   MEDLINE=21371909; PubMed=11479226;
RA   Yang D., Holt G.E., Velders M.P., Kwon E.D., Kast W.M.;
RT   "Murine six-transmembrane epithelial antigen of the prostate, prostate
RT   stem cell antigen, and prostate-specific membrane antigen: prostate-
RT   specific cell-surface antigens highly expressed in prostate cancer of
RT   transgenic adenocarcinoma mouse prostate mice.";
RL   Cancer Res. 61:5857-5860(2001).
DR   EMBL; AF297098; AAK83126.1; -.
DR   MGD; MGI:1917608; Steap.
KW   Transmembrane.
SQ   SEQUENCE   339 AA;  39318 MW;  4B26A71FF559E84F CRC64;

  Query Match             29.9%;  Score 704;  DB 11;  Length 339;
  Best Local Similarity   52.9%;  Pred. No. 3.8e-48;
  Matches  128;  Conservative   51;  Mismatches   63;  Indels    0;  Gaps    0;

Qy     209 FTLWRGPVVVAISLATFFFLYSFVRDVIHPYARNQQSDFYKIPIEIVNKTLPIVAITLLS 268
Db      68 FPNWRLPVKVAAIISSLTFLYTLLREIIYPLVTSREQYFYKIPILVINKVLPMVAITLLA 127

Qy     269 LVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQCRKQLGLLSFFFAMVHVAYSLCLPMRRS 328
Db     128 LVYLPGELAAVVQLRNGTKYKKFPPWLDRWMLARKQFGLLSFFFAVLYAVYSLSYPMRRS 187

Qy     329 ERYLFLNMAYQQVHANIENSWNEEEVWRIEMYISFGIMSLGLLSLLAVTSIPSVSNALNW 388
Db     188 YRYKLLNWAYKQVQQNKEDAWVEHDVWRMEIYVSLGIVGLAILALLAVTSIPSVSDSLTW 247

Qy     389 REFSFIQSTLGYVALLISTFHVLIYGWKRAFEEEYYRFYTPPNFVLALVLPSIVILDLLQ 448
Db     248 REFHYIQSKLGIVSLLLGTVHALVFAWNKWVDVSQFVWYMPPTFMIAVFLPTLVLICKIA 307

Qy     449 LC 450
Db     308 LC 309


RESULT 13
Q9H5R1
ID   Q9H5R1      PRELIMINARY;      PRT;   283 AA.
AC   Q9H5R1;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CDNA: FLJ23153 fis, clone LNG09441.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LUNG;
RA   Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,
RA   Nakajima Y., Mizuno T., Morinaga M., Tanigami A., Fujiwara T., Ono T.,
RA   Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y., Ota T., Suzuki Y.,
RA   Obayashi M., Nishi T., Shibahara T., Tanaka T., Nakamura Y.,
RA   Isogai T., Sugano S.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK026806; BAB15559.1; -.
DR   InterPro; IPR000585; Hemopexin.
DR   PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
SQ   SEQUENCE   283 AA;  31814 MW;  DDD330DA7C607D06 CRC64;

  Query Match             26.7%;  Score 628;  DB 4;  Length 283;
  Best Local Similarity   43.2%;  Pred. No. 3.5e-42;
  Matches  123;  Conservative   56;  Mismatches   80;  Indels   26;  Gaps    5;

Qy      13 LSETC---LPNGINGIKDARKVTVGVIGSGDFAKSLTIRLIRCGYHVVIGSRNPKFASEF 69
Db       1 MEKTCIDALPLTMN--SSEKQETVCIFGTGDFGRSLGLKMLQCGYSVVFGSRNPQ-KTTL 57

Qy      70 FPHVVDVTHHEDALTKTNIIFVAIHREHYTSLWDLRHLLVGKILIDVSNNMRINQYPESN 129
Db      58 LPSGAEVLSYSEAAKKSDIIIIAIHREHYDFLTELTEVLNGKILVDISNNLKINQYPESN 117

Qy     130 AEYLASLFPDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIELARQLNFIP 189
Db     118 AEYLAHLVPGAHVVKAFNTISAWALQSGALDASRQVFVCGNDSKAKQRVMDIVRNLGLTP 177

Qy     190 IDLGSLSSAREIENLPLRLFTLWRGPVVVAISLATFFFLYSFVRDVIHPYARNQQSDFYK 249
Db     178 MDQGSLMAAKEIEKYPLQLFPMWRFPFYLSAVLCVFLFFYCVIRDVIYPYVYEKKDNTFR 237

Qy     250 IPIEIVNKTLPIVAITLLSLVYLAGLLAAAYQLYYGTKYRRFPPW 294
Db     238 MAISIPNRIFPI--------------TAPYTACFG-----LPPW 262


RESULT 14
Q9Y6U5
ID   Q9Y6U5      PRELIMINARY;      PRT;   264 AA.
AC   Q9Y6U5;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   WUGSC:H_RG087E15.1 protein (Fragment).
GN   WUGSC:H_RG087E15.1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99063792; PubMed=9847074;
RA   Sulston J.E., Waterston R.;
RT   "Toward a complete human genome sequence.";
```

```
RL   Genome Res. 8:1097-1108(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Strong C., Layman D., Graves T., Strowmatt C.;
RT   "The sequence of Homo sapiens BAC clone CTB-87E15.";
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Waterston R.;
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC005061; AAD43182.1; -.
FT   NON_TER       1      1
FT   NON_TER     264    264
SQ   SEQUENCE   264 AA;  31203 MW;  35C9483003557E72 CRC64;

  Query Match             24.5%;  Score 577;  DB 4;  Length 264;
  Best Local Similarity   56.4%;  Pred. No. 3.7e-38;
  Matches  106;  Conservative   36;  Mismatches   46;  Indels    0;  Gaps    0;

Qy     208 LFTLWRGPVVVAISLATFFFLYSFVRDVIHPYARNQQSDFYKIPIEIVNKTLPIVAITLL 267
           ||  |  |: :|  :|:  |||: :|:|||| | : |  |||||| ::|| ||:|:|||
Db      77 LFPQWHLPIKIAAVMASLTFLYTLLREVIHPLATSHQQYFYKIPILVINKVLPMVSITLL 136

Qy     268 SLVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQCRKQLGLLSFFFAMVHVAYSLCLPMRR 327
           :|||| |::||  |:: ||||::|| ||: |:  ||| |||| |||::|  |:|   |||
Db     137 ALVYLPGVIAAIVQVHNGTKYKKFPHWLDKWMLTRKQFGLLSLFFAVLHAIYTLSYAMRR 196

Qy     328 SERYLFLNMAYQQVHANIENSWNEEEVWRIEMYISFGIMSLGLLSLLAVTSIPSVSNALN 387
           | ||  || |||||  | |::| | :|||:|:|:| ||: | :|:||||||||||::|
Db     197 SYRYKLLNWAYQQVQQNKEDAWIEHDVWRMEIYVSLGIVGLAILALLAVTSIPSVSDSLT 256

Qy     388 WREFSFIQ 395
           |||| :||
Db     257 WREFHYIQ 264


RESULT 15
Q8WWB0
ID   Q8WWB0      PRELIMINARY;      PRT;   283 AA.
AC   Q8WWB0;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Similar to likely ortholog of mouse tumor necrosis-alpha-induced
DE   adipose-related protein.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LUNG;
RA   Strausberg R.;
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC020600; AAH20600.1; -.
SQ   SEQUENCE   283 AA;  31310 MW;  05110EAC5D032317 CRC64;

  Query Match             23.0%;  Score 541;  DB 4;  Length 283;
  Best Local Similarity   28.5%;  Pred. No. 3e-35;
  Matches  124;  Conservative   50;  Mismatches   79;  Indels  182;  Gaps    5;

Qy      13 LSETC---LPNGINGIKDARKVTVGVIGSGDFAKSLTIRLIRCGYHVVIGSRNPKFASEF 69
           : :||   ||  :|     :: || : |:||| :|| :::::||| || ||||||  :
Db       1 MEKTCIDALPLTMN--SSEKQETVCIFGTGDFGRSLGLKMLQCGYSVVFGSRNPQ-KTTL 57

Qy      70 FPHVVDVTHHEDALTKTNIIFVAIHREHYTSLWDLRHLLVGKILIDVSNNMRINQYPESN 129
             |   :|  : :|  |: || :|||||||  | :|  :| ||||:|:|||::||||||||
Db      58 LPSGAEVLSYSEAAKKSGIIIIAIHREHYDFLTELTEVLNGKILVDISNNLKINQYPESN 117

Qy     130 AEYLASLFPDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIELARQLNFIP 189
           |||||  | | : :|| || :|||||  |  |||||
Db     118 AEYLAHLVPGAHVVKAFNTISAWALQSGALDASRQA---------------------- 153

Qy     190 IDLGSLSSAREIENLPLRLFTLWRGPVVVAISLATFFFLYSFVRDVIHPYARNQQSDFYK 249

Db     154 ------------------------------------------------------------ 153

Qy     250 IPIEIVNKTLPIVAITLLSLVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQCRKQLGLLS 309
               |: |
Db     154 ----ILKK---------------------------------------------------- 157

Qy     310 FFFAMVHVAYSLCLPMRRSERYLFLNMAYQQVHANIENSWNEEEVWRIEMYISFGIMSLG 369
                                                || ::    |  : |:: ||:
Db     158 ----------------------------------ENPFSTSSAWLSDSYVALGILGFF 181

Qy     370 LLSLLAVTSIPSVSNALNWREFSFIQSTLGYVALLISTFHVLIYGWKRAFEEEYYRFYTP 429
           |  || :||:||||||:||||| |:|| |||: |:: | | |:|| ||       |:| |
Db     182 LFVLLGITSLPSVSNAVNWREFRFVQSKLGYLTLILCTAHTLVYGGKRFLSPSNLRWYLP 241

Qy     430 PNFVLALVLPSIVIL 444
             :|| |::|  |::
Db     242 AAYVLGLIIPCTVLV 256


Search completed: January 17, 2003, 19:00:58
Job time : 41 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: January 17, 2003, 19:00:21 ; Search time 16 Seconds
(without alignments)
564.053 Million cell updates/sec

Title: US-09-455-486-6
Perfect score: 2351
Sequence: 1 MESISMMGSPKSLSETCLPN.....ALVLPSIVILDLLQLCRYPD 454

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 120991 seqs, 19878514 residues

Total number of hits satisfying chosen parameters: 120991

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2294	97.6	490	9	US-09-802-520-1	Sequence 1, Appli
2	1269	54.0	488	10	US-09-747-835A-13	Sequence 13, Appl
3	1088	46.3	459	9	US-09-965-529-21	Sequence 21, Appl
4	844	35.9	179	10	US-09-864-761-37654	Sequence 37654, A
5	768	32.7	237	10	US-09-747-835A-15	Sequence 15, Appl
6	736	31.3	141	10	US-09-963-896-1	Sequence 1, Appli
7	717	30.5	267	10	US-09-747-835A-50	Sequence 50, Appl
8	717	30.5	267	10	US-09-747-835A-51	Sequence 51, Appl
9	717	30.5	339	9	US-10-012-896-879	Sequence 879, App
10	717	30.5	339	9	US-09-802-520-11	Sequence 11, Appl
11	717	30.5	339	9	US-09-895-793-879	Sequence 879, App
12	717	30.5	339	9	US-09-895-814-879	Sequence 879, App
13	717	30.5	339	10	US-09-759-143-879	Sequence 879, App
14	717	30.5	339	10	US-09-780-669-879	Sequence 879, App
15	717	30.5	339	10	US-09-822-827-879	Sequence 879, App
16	324	13.8	132	10	US-09-789-561-89	Sequence 89, Appl
17	202	8.6	104	10	US-09-864-761-47234	Sequence 47234, A
18	196	8.3	38	10	US-09-864-761-38670	Sequence 38670, A
19	171.5	7.3	208	9	US-09-738-626-6627	Sequence 6627, Ap
20	107.5	4.6	695	10	US-09-804-626-8	Sequence 8, Appli
21	104	4.4	396	10	US-09-815-242-10164	Sequence 10164, A
22	101	4.3	476	10	US-09-774-414-3	Sequence 3, Appli
23	101	4.3	735	10	US-09-815-242-12089	Sequence 12089, A
24	99	4.2	365	9	US-10-109-533A-2	Sequence 2, Appli
25	97.5	4.1	516	10	US-09-804-626-4	Sequence 4, Appli
26	97	4.1	675	10	US-09-877-804-7	Sequence 7, Appli
27	97	4.1	692	10	US-09-877-804-6	Sequence 6, Appli
28	96	4.1	299	10	US-09-903-456-61	Sequence 61, Appl
29	95	4.0	224	9	US-09-738-626-4894	Sequence 4894, Ap
30	94	4.0	284	10	US-09-903-456-83	Sequence 83, Appl
31	94	4.0	292	10	US-09-903-456-35	Sequence 35, Appl
32	94	4.0	293	10	US-09-903-456-45	Sequence 45, Appl
33	91	3.9	485	9	US-10-055-025-2	Sequence 2, Appli
34	90.5	3.8	465	10	US-09-915-181A-8	Sequence 8, Appli
35	90.5	3.8	1617	9	US-10-090-453A-2	Sequence 2, Appli
36	89.5	3.8	968	10	US-09-753-008-7	Sequence 7, Appli
37	89	3.8	1624	12	US-10-090-454-2	Sequence 2, Appli
38	88.5	3.8	675	10	US-09-733-630-2	Sequence 2, Appli
39	88	3.7	387	8	US-08-945-749-8	Sequence 8, Appli
40	87.5	3.7	299	10	US-09-903-456-66	Sequence 66, Appl
41	87.5	3.7	675	9	US-09-928-530-2	Sequence 2, Appli
42	87	3.7	348	10	US-09-877-804-8	Sequence 8, Appli
43	87	3.7	467	10	US-09-991-212A-3	Sequence 3, Appli
44	87	3.7	801	10	US-09-815-242-5518	Sequence 5518, Ap
45	87	3.7	801	10	US-09-815-242-12519	Sequence 12519, A

ALIGNMENTS

RESULT 1
US-09-802-520-1
; Sequence 1, Application US/09802520
; Publication No. US20020187472A1
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Faris, Mary
; APPLICANT: Chen, Huei-Mei
; APPLICANT: Ison, Craig H.
; TITLE OF INVENTION: STEAP-RELATED PROTEIN
; FILE REFERENCE: PC-0037 US
; CURRENT APPLICATION NUMBER: US/09/802,520
; CURRENT FILING DATE: 2001-03-09
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PERL Program
; SEQ ID NO 1
; LENGTH: 490
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020187472A1 7492448CD1
US-09-802-520-1

Query Match 97.6%; Score 2294; DB 9; Length 490;
Best Local Similarity 100.0%; Pred. No. 1.1e-205;
Matches 444; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MESISMMGSPKSLSETCLPNGINGIKDARKVTVGVIGSGDFAKSLTIRLIRCGYHVVIGS 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MESISMMGSPKSLSETCLPNGINGIKDARKVTVGVIGSGDFAKSLTIRLIRCGYHVVIGS 60

Qy      61 RNPKFASEFFPHVVDVTHHEDALTKTNIIFVAIHREHYTSLWDLRHLLVGKILIDVSNNM 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 RNPKFASEFFPHVVDVTHHEDALTKTNIIFVAIHREHYTSLWDLRHLLVGKILIDVSNNM 120

Qy     121 RINQYPESNAEYLASLFPDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIE 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 RINQYPESNAEYLASLFPDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIE 180

Qy     181 LARQLNFIPIDLGSLSSAREIENLPLRLFTLWRGPVVVAISLATFFFLYSFVRDVIHPYA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    181 LARQLNFIPIDLGSLSSAREIENLPLRLFTLWRGPVVVAISLATFFFLYSFVRDVIHPYA 240

Qy    241 RNQQSDFYKIPIEIVNKTLPIVAITLLSLVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQ 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 RNQQSDFYKIPIEIVNKTLPIVAITLLSLVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQ 300

Qy    301 CRKQLGLLSFFFAMVHVAYSLCLPMRRSERYLFLNMAYQQVHANIENSWNEEEVWRIEMY 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CRKQLGLLSFFFAMVHVAYSLCLPMRRSERYLFLNMAYQQVHANIENSWNEEEVWRIEMY 360

Qy    361 ISFGIMSLGLLSLLAVTSIPSVSNALNWREFSFIQSTLGYVALLISTFHVLIYGWKRAFE 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ISFGIMSLGLLSLLAVTSIPSVSNALNWREFSFIQSTLGYVALLISTFHVLIYGWKRAFE 420

Qy    421 EEYYRFYTPPNFVLALVLPSIVIL 444
          ||||||||||||||||||||||||
Db    421 EEYYRFYTPPNFVLALVLPSIVIL 444


RESULT 2
US-09-747-835A-13
; Sequence 13, Application US/09747835A
; Patent No. US20020146692A1
; GENERAL INFORMATION:
;  APPLICANT: Yamazaki, Victoria
;  APPLICANT:  Tang, Y. Tom
;  APPLICANT:  Liu, Chenghua
;  APPLICANT:  Zhou, Ping
;  APPLICANT:  Wang, Dunrui
;  APPLICANT:  Zhang, Jie
;  APPLICANT:  Ren, Feiyan
;  APPLICANT:  Asundi, Vinod
;  APPLICANT:  Drmanac, Radoje T
;  TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO G PROTEIN-COUPLED RECEPTOR-LIKE
;  TITLE OF INVENTION:  LIKE) POLYPEPTIDES AND POLYNUCLEOTIDES
;  FILE REFERENCE: HYS-37CIP
;  CURRENT APPLICATION NUMBER: US/09/747,835A
;  CURRENT FILING DATE:  2002-03-08
;  PRIOR APPLICATION NUMBER: US 09/729,739
;  PRIOR FILING DATE: 2000-12-04
;  PRIOR APPLICATION NUMBER: US 09/653,450
;  PRIOR FILING DATE: 2000-08-31
;  PRIOR APPLICATION NUMBER: US 09/620,312
;  PRIOR FILING DATE: 2000-07-19
;  PRIOR APPLICATION NUMBER: US 09/598,042
;  PRIOR FILING DATE: 2000-06-20
;  PRIOR APPLICATION NUMBER: US 09/552,317
;  PRIOR FILING DATE: 2000-04-25
;  PRIOR APPLICATION NUMBER: US 09/488,725
;  PRIOR FILING DATE: 2000-01-21
;  NUMBER OF SEQ ID NOS: 63
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 13
;   LENGTH: 488
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-747-835A-13

  Query Match           54.0%;  Score 1269;  DB 10;  Length 488;
  Best Local Similarity  54.3%;  Pred. No. 2.8e-110;
  Matches  241;  Conservative   83;  Mismatches  108;  Indels   12;  Gaps    4;

Qy      3 SISMMGSPKSLSETCLPNGINGIKDARKVTVGVIGSGDFAKSLTIRLIRCGYHVVIGSRN 62
          |: :: |  ||::  :|:      :| |  ||::|||||:||  ||:  |: ||:|||
Db     11 SLHLVDSDSSLAK--VPD------EAPK--VGILGSGDFARSLATRLVGSGFKVVVGSRN 60

Qy     63 PKFASEFFPHVVDVTHHEDALTKTNIIFVAIHREHYTSLWDLRHLLVGKILIDVSN--NM 120
          ||  :  :|    ||  |:|::   :||||: ||||:||  |   | ||||:||||
Db     61 PKRTARLYPSAAQVTFQEEAVSSPEVIFVAVFREHYSSLCSLSDQLAGKILVDVSNPTEQ 120

Qy    121 RINQYPESNAEYLASLFPDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIE 180
          |: ||||||||||||   :|| |||:||| || ||:| :||| || :  :|:: | |
Db    121 EHLQHRESNAEYLASLFPTCTVVKAFNVISAWTLQAGPRDGNRQVPICGDQPEAKRAVSE 180


Qy    181 LARQLNFIPIDLGSLSSAREIENLPLRLFTLWRGPVVVAISLATFFFLYSFVRDVIHPYA 240
          :|  : |:|:|:|||:|| |:| :||||   |: | ::|: |   |: |:|||||: ||
Db    181 MALAMGFMPVDMGSLASAWEVEAMPLRLLPAWKVPTLLALGLFVCFYAYNFVRDVLQPYV 240

Qy    241 RNQQSDFYKIPIEIVNKTLPIVAITLLSLVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQ 300
          :  |: |:|:|: :|| ||| ||  ||||||| |:|||| ||  ||||:||| ||: |||
Db    241 QESQNKFFKLPVSVVNTTLPCVAYVLLSLVYLPGVLAAALQLRRGTKYQRFPDWLDHWLQ 300

Qy    301 CRKQLGLLSFFFAMVHVAYSLCLPMRRSERYLFLNMAYQQVHANIENSWNEEEVWRIEMY 360
          |||:|||||| | :|  || |||:||: ||  :|:| :|| ||  : | ||||||:|:|
Db    301 HRKQIGLLSFFCAALHALYSFCLPLRRAHRYDLVNLAVKQVLANKSHLWVEEEVWRMEIY 360

Qy    361 ISFGIMSLGLLSLLAVTSIPSVSNALNWREFSFIQSTLGYVALLISTFHVLIYGWKRAFE 420
          :| |:::|| ||||||||:||::|:||||||||:||:||:|||::|| | | ||| ||||
Db    361 LSLGVLALGTLSLLAVTSLPSIANSLNWREFSFVQSSLGFVALVLSTLHTLTYGWTRAFE 420

Qy    421 EEYYRFYTPPNFVLALVLPSIVIL 444
          |  |:|| || | | |::| :|||
Db    421 ESRYKFYLPPTFTLTLLVPCVVIL 444


RESULT 3
US-09-965-529-21
; Sequence 21, Application US/09965529
; Publication No. US20020182671A1
; GENERAL INFORMATION:
;  APPLICANT: LAL, Preeti
;  APPLICANT:  YUE, Henry
;  APPLICANT:  TANG, Y. Tom
;  APPLICANT:  BANDMAN, Olga
;  APPLICANT:  BURFORD, Neil
;  APPLICANT:  AZIMZAI, Yalda
;  APPLICANT:  BAUGHN, Mariah R.
;  APPLICANT:  LU, Dyung Aina M.
;  APPLICANT:  PATTERSON, Chandra
;  TITLE OF INVENTION: MEMBRANE ASSOCIATED PROTEINS
;  FILE REFERENCE: PF-0731 USA
;  CURRENT APPLICATION NUMBER: US/09/965,529
;  CURRENT FILING DATE:  2001-09-26
;  PRIOR APPLICATION NUMBER: 60/149,641; 60/164,203; PCT/US00/22315
;  PRIOR FILING DATE: 1999-08-17; 1999-11-09; 2000-08-14
;  NUMBER OF SEQ ID NOS: 74
;  SOFTWARE: PERL Program
; SEQ ID NO 21
;   LENGTH: 459
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20020182671A1 1859305CD1
US-09-965-529-21

  Query Match           46.3%;  Score 1088;  DB 9;  Length 459;
  Best Local Similarity  46.9%;  Pred. No. 1.8e-93;
  Matches  204;  Conservative   89;  Mismatches  136;  Indels    6;  Gaps    3;

Qy     13 LSETC---LPNGINGIKDARKVTVGVIGSGDFAKSLTIRLIRCGYHVVIGSRNPKFASEF 69
          : :||   ||  :|     :: || : |:||| :|| :::::||| || |||||:  :
Db      1 MEKTCIDALPLTMN--SSEKQETVCIFGTGDFGRSLGLKMLQCGYSVVFGSRNPQ-KTTL 57

Qy     70 FPHVVDVTHHEDALTKTNIIFVAIHREHYTSLWDLRHLLVGKILIDVSNNMRINQYPESN 129
          |   :|  : :|  |::|| :|||||||  | :|  :| |||||:|:|||::|||||||
Db     58 LPSGAEVLSYSEAAKKSDIIIIAIHREHYDFLTELTEVLNGKILVDISNNLKINQYPESN 117

Qy    130 AEYLASLFPDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIELARQLNFIP 189
          ||||| | | : :|| || :|||||| |  ||||||::| |: :|:|:|::: | |  ||
Db    118 AEYLAHLVPGAHVVKAFNTISAWALQSGALDASRQVFVCGNDSKAKQRVMDIVRNLGLTP 177

Qy    190 IDLGSLSSAREIENLPLRLFTLWRGPVVVAISLATFFFLYSFVRDVIHPYARNQQSDFYK 249
          :| ||| :|:|||  ||:|| :|| |  ::  |  | | |  :||||:||   :: : ::
Db    178 MDQGSLMAAKEIEKYPLQLFPMWRFPFYLSAVLCVFLFFYCVIRDVIYPYVYEKKDNTFR 237
```

```
Qy     250 IPIEIVNKTLPIVAITLLSLVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQCRKQLGLLS 309
           : | | |:  || |:|||:|||| |::||  ||| |||||||| ||: |: |||||||::
Db     238 MAISIPNRIFPITALTLLALVYLPGVIAAILQLYRGTKYRRFPDWLDHWMLCRKQLGLVA 297

Qy     310 FFFAMVHVAYSLCLPMRRSERYLFLNMAYQQVHANIENSWNEEEVWRIEMYISFGIMSLG 369
             || :|| |:| :|:|   |:   |:   |     || ::    |  : |:: ||:
Db     298 LGFAFLHVLYTLVIPIRYYVRWRLGNLTVTQAILKKENPFSTSSAWLSDSYVALGILGFF 357

Qy     370 LLSLLAVTSIPSVSNALNWREFSFIQSTLGYVALLISTFHVLIYGWKRAFEEEYYRFYTP 429
           |  || :||:|||||:|||| |:|| ||| |:: | | |:|| ||       |:| |
Db     358 LFVLLGITSLPSVSNAVNWREFRFVQSKLGYLTLILCTAHTLVYGGKRFLSPSNLRWYLP 417

Qy     430 PNFVLALVLPSIVIL 444
           :|| |::|  |::
Db     418 AAYVLGLIIPCTVLV 432


RESULT 4
US-09-864-761-37654
; Sequence 37654, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 37654
;   LENGTH: 179
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC002064.1
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.8
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.2
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.9
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.1
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.1
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.4
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.7
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.6
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.3
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.2
;   OTHER INFORMATION: SWISSPROT HIT: P23523, EVALUE 8.00e-01
;   OTHER INFORMATION: EST_HUMAN HIT: BE783685.1, EVALUE 9.00e-37
US-09-864-761-37654

  Query Match             35.9%;  Score 844;  DB 10;  Length 179;
  Best Local Similarity   98.8%;  Pred. No. 2.5e-71;
  Matches  165;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 MESISMMGSPKSLSETCLPNGINGIKDARKVTVGVIGSGDFAKSLTIRLIRCGYHVVIGS 60
           |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db      12 MESISMMGSPKSLSETFLPNGINGIKDARKVTVGVIGSGDFAKSLTIRLIRCGYHVVIGS 71

Qy      61 RNPKFASEFFPHVVDVTHHEDALTKTNIIFVAIHREHYTSLWDLRHLLVGKILIDVSNNM 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72 RNPKFASEFFPHVVDVTHHEDALTKTNIIFVAIHREHYTSLWDLRHLLVGKILIDVSNNM 131

Qy     121 RINQYPESNAEYLASLFPDSLIVKGFNVVSAWALQLGPKDASRQVYI 167
           ||||||||||||||||||||||||||||||||||||||||||||||| |
Db     132 RINQYPESNAEYLASLFPDSLIVKGFNVVSAWALQLGPKDASRQVCI 178


RESULT 5
US-09-747-835A-15
; Sequence 15, Application US/09747835A
; Patent No. US20020146692A1
; GENERAL INFORMATION:
;  APPLICANT: Yamazaki, Victoria
;  APPLICANT:  Tang, Y. Tom
;  APPLICANT:  Liu, Chenghua
;  APPLICANT:  Zhou, Ping
;  APPLICANT:  Wang, Dunrui
;  APPLICANT:  Zhang, Jie
;  APPLICANT:  Ren, Feiyan
;  APPLICANT:  Asundi, Vinod
;  APPLICANT:  Drmanac, Radoje T
;  TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO G PROTEIN-COUPLED RECEPTOR-L
;  TITLE OF INVENTION:  LIKE) POLYPEPTIDES AND POLYNUCLEOTIDES
;  FILE REFERENCE: HYS-37CIP
;  CURRENT APPLICATION NUMBER: US/09/747,835A
;  CURRENT FILING DATE:  2002-03-08
;  PRIOR APPLICATION NUMBER: US 09/729,739
;  PRIOR FILING DATE: 2000-12-04
;  PRIOR APPLICATION NUMBER: US 09/653,450
;  PRIOR FILING DATE: 2000-08-31
;  PRIOR APPLICATION NUMBER: US 09/620,312
;  PRIOR FILING DATE: 2000-07-19
;  PRIOR APPLICATION NUMBER: US 09/598,042
;  PRIOR FILING DATE: 2000-06-20
;  PRIOR APPLICATION NUMBER: US 09/552,317
;  PRIOR FILING DATE: 2000-04-25
;  PRIOR APPLICATION NUMBER: US 09/488,725
;  PRIOR FILING DATE: 2000-01-21
;  NUMBER OF SEQ ID NOS: 63
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 15
;   LENGTH: 237
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-747-835A-15

  Query Match             32.7%;  Score 768;  DB 10;  Length 237;
```

```
  Best Local Similarity   60.0%;  Pred. No. 4.4e-64;
  Matches  138;  Conservative   42;  Mismatches   50;  Indels    0;  Gaps    0;

Qy     215 PVVVAISLATFFFLYSFVRDVIHPYARNQQSDFYKIPIEIVNKTLPIVAITLLSLVYLAG 274
           | ::|: |   |: |:|||||: || :  |: |:|:|: :|| ||| ||  ||||||| |
Db       2 PTLLALGLFVCFYAYNFVRDVLQPYVQESQNKFFKLPVSVVNTTLPCVAYVLLSLVYLPG 61

Qy     275 LLAAAYQLYYGTKYRRFPPWLETWLQCRKQLGLLSFFFAMVHVAYSLCLPMRRSERYLFL 334
           :|||| ||  |||||||| ||: ||| |||:|||||| | :|  || |||:||: ||  :
Db      62 VLAAALQLRRGTKYQRFPDWLDHWLQHRKQIGLLSFFCAALHALYSFCLPLRRAHRYDLV 121

Qy     335 NMAYQQVHANIENSWNEEEVWRIEMYISFGIMSLGLLSLLAVTSIPSVSNALNWREFSFI 394
           |:| :|| ||  : | |||||:|:|:| |:::|| |||||||:||::|:|||||||:
Db     122 NLAVKQVLANKSHLWVEEEVWRMEIYLSLGVLALGTLSLLAVTSLPSIANSLNWREFSFV 181

Qy     395 QSTLGYVALLISTFHVLIYGWKRAFEEEYYRFYTPPNFVLALVLPSIVIL 444
           ||:||:|||::|| | | ||| ||||| |:|| || | | |::| :|||
Db     182 QSSLGFVALVLSTLHTLTYGWTRAFEESRYKFYLPPTFTLTLLVPCVVIL 231


RESULT 6
US-09-963-896-1
; Sequence 1, Application US/09963896
; Patent No. US20020102585A1
;   GENERAL INFORMATION:
;        APPLICANT: Lal, Preeti
;                   Guegler, Karl J.
;                   Corley, Neil C.
;        TITLE OF INVENTION: PROSTATE GROWTH-ASSOCIATED MEMBRANE PROTEINS
;        NUMBER OF SEQUENCES: 7
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
;             STREET: 3174 PORTER DRIVE
;             CITY: PALO ALTO
;             STATE: CALIFORNIA
;             COUNTRY: USA
;             ZIP: 94304
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/963,896
;             FILING DATE: 26-Sep-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 09/397,558
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: CERRONE, MICHAEL C.
;             REGISTRATION NUMBER: 39,132
;             REFERENCE/DOCKET NUMBER: PF-0527 US
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (650) 855-0555
;             TELEFAX: (650) 845-4166
;   INFORMATION FOR SEQ ID NO: 1:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 141 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        IMMEDIATE SOURCE:
;             LIBRARY: PROSTUT10
;             CLONE: 1691243
;        SEQUENCE DESCRIPTION: SEQ ID NO: 1 :
US-09-963-896-1

  Query Match           31.3%;  Score 736;  DB 10;  Length 141;
  Best Local Similarity  100.0%;  Pred. No. 2e-61;
  Matches  141;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     314 MVHVAYSLCLPMRRSERYLFLNMAYQQVHANIENSWNEEEVWRIEMYISFGIMSLGLLSL 373
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MVHVAYSLCLPMRRSERYLFLNMAYQQVHANIENSWNEEEVWRIEMYISFGIMSLGLLSL 60

Qy     374 LAVTSIPSVSNALNWREFSFIQSTLGYVALLISTFHVLIYGWKRAFEEEYYRFYTPPNFV 433
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 LAVTSIPSVSNALNWREFSFIQSTLGYVALLISTFHVLIYGWKRAFEEEYYRFYTPPNFV 120

Qy     434 LALVLPSIVILDLLQLCRYPD 454
           |||||||||||||||||||||
Db     121 LALVLPSIVILDLLQLCRYPD 141


RESULT 7
US-09-747-835A-50
; Sequence 50, Application US/09747835A
; Patent No. US20020146692A1
; GENERAL INFORMATION:
;  APPLICANT: Yamazaki, Victoria
;  APPLICANT:  Tang, Y. Tom
;  APPLICANT:  Liu, Chenghua
;  APPLICANT:  Zhou, Ping
;  APPLICANT:  Wang, Dunrui
;  APPLICANT:  Zhang, Jie
;  APPLICANT:  Ren, Feiyan
;  APPLICANT:  Asundi, Vinod
;  APPLICANT:  Drmanac, Radoje T
;  TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO G PROTEIN-COUPLED RECEPTOR-L
;  TITLE OF INVENTION:  LIKE) POLYPEPTIDES AND POLYNUCLEOTIDES
;  FILE REFERENCE: HYS-37CIP
;  CURRENT APPLICATION NUMBER: US/09/747,835A
;  CURRENT FILING DATE:  2002-03-08
;  PRIOR APPLICATION NUMBER: US 09/729,739
;  PRIOR FILING DATE: 2000-12-04
;  PRIOR APPLICATION NUMBER: US 09/653,450
;  PRIOR FILING DATE: 2000-08-31
;  PRIOR APPLICATION NUMBER: US 09/620,312
;  PRIOR FILING DATE: 2000-07-19
;  PRIOR APPLICATION NUMBER: US 09/598,042
;  PRIOR FILING DATE: 2000-06-20
;  PRIOR APPLICATION NUMBER: US 09/552,317
;  PRIOR FILING DATE: 2000-04-25
;  PRIOR APPLICATION NUMBER: US 09/488,725
;  PRIOR FILING DATE: 2000-01-21
;  NUMBER OF SEQ ID NOS: 63
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 50
;   LENGTH: 267
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-747-835A-50

  Query Match           30.5%;  Score 717;  DB 10;  Length 267;
  Best Local Similarity  54.9%;  Pred. No. 2.9e-59;
  Matches  130;  Conservative   48;  Mismatches   59;  Indels    0;  Gaps    0;

Qy     208 LFTLWRGPVVVAISLATFFFLYSFVRDVIHPYARNQQSDFYKIPIEIVNKTLPIVAITLL 267
           ||  |  |: :|  :|:  |||: :|:|||| | : |  |||||| ::|| ||:|:|||
Db       1 LFPQWHLPIKIAAIIASLTFLYTLLREVIHPLATSHQQYFYKIPILVINKVLPMVSITLL 60

Qy     268 SLVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQCRKQLGLLSFFFAMVHVAYSLCLPMRR 327
           :|||| |::||  ||: ||||::|| ||: |:  ||| ||||||||::|  |||  ||||
Db      61 ALVYLPGVIAAIVQLHNGTKYKKFPHWLDKWMLTRKQFGLLSFFFAVLHAIYSLSYPMRR 120

Qy     328 SERYLFLNMAYQQVHANIENSWNEEEVWRIEMYISFGIMSLGLLSLLAVTSIPSVSNALN 387
           | ||  || ||||||  | |::| | :|||:|:|:| ||: | :|:|||||||||::|
Db     121 SYRYKLLNWAYQQVQQNKEDAWIEHDVWRMEIYVSLGIVGLAILALLAVTSIPSVSDSLT 180

Qy     388 WREFSFIQSTLGYVALLISTFHVLIYGWKRAFEEEYYRFYTPPNFVLALVLPSIVIL 444
           |||| :||| || |:||: | | ||: | :  : : : :|||| |::|: || :|::
Db     181 WREFHYIQSKLGIVSLLLGTIHALIFAWNKWIDIKQFVWYTPPTFMIAVFLPIVVLI 237


RESULT 8
```

```
US-09-747-835A-51
; Sequence 51, Application US/09747835A
; Patent No. US20020146692A1
; GENERAL INFORMATION:
;  APPLICANT: Yamazaki, Victoria
;  APPLICANT:  Tang, Y. Tom
;  APPLICANT:  Liu, Chenghua
;  APPLICANT:  Zhou, Ping
;  APPLICANT:  Wang, Dunrui
;  APPLICANT:  Zhang, Jie
;  APPLICANT:  Ren, Feiyan
;  APPLICANT:  Asundi, Vinod
;  APPLICANT:  Drmanac, Radoje T
;  TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO G PROTEIN-COUPLED RECEPTOR-LIKE
;  TITLE OF INVENTION:  LIKE) POLYPEPTIDES AND POLYNUCLEOTIDES
;  FILE REFERENCE: HYS-37CIP
;  CURRENT APPLICATION NUMBER: US/09/747,835A
;  CURRENT FILING DATE:  2002-03-08
;  PRIOR APPLICATION NUMBER: US 09/729,739
;  PRIOR FILING DATE: 2000-12-04
;  PRIOR APPLICATION NUMBER: US 09/653,450
;  PRIOR FILING DATE: 2000-08-31
;  PRIOR APPLICATION NUMBER: US 09/620,312
;  PRIOR FILING DATE: 2000-07-19
;  PRIOR APPLICATION NUMBER: US 09/598,042
;  PRIOR FILING DATE: 2000-06-20
;  PRIOR APPLICATION NUMBER: US 09/552,317
;  PRIOR FILING DATE: 2000-04-25
;  PRIOR APPLICATION NUMBER: US 09/488,725
;  PRIOR FILING DATE: 2000-01-21
;  NUMBER OF SEQ ID NOS: 63
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 51
;   LENGTH: 267
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-747-835A-51

  Query Match             30.5%;  Score 717;  DB 10;  Length 267;
  Best Local Similarity   54.9%;  Pred. No. 2.9e-59;
  Matches  130;  Conservative   48;  Mismatches   59;  Indels    0;  Gaps    0;

Qy    208 LFTLWRGPVVVAISLATFFFLYSFVRDVIHPYARNQQSDFYKIPIEIVNKTLPIVAITLL 267
          ||  |  |: :|  :|:  ||| :|:|||| | : |  |||||| ::|| ||:|:|||| 
Db      1 LFPQWHLPIKIAAIIASLTFLYTLLREVIHPLATSHQQYFYKIPILVINKVLPMVSITLL 60

Qy    268 SLVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQCRKQLGLLSFFFAMVHVAYSLCLPMRR 327
          :|||| |::||  ||: ||||::|| ||: |:  ||| |||||||::|  |||  |||| 
Db     61 ALVYLPGVIAAIVQLHNGTKYKKFPHWLDKWMLTRKQFGLLSFFFAVLHAIYSLSYPMRR 120

Qy    328 SERYLFLNMAYQQVHANIENSWNEEEVWRIEMYISFGIMSLGLLSLLAVTSIPSVSNALN 387
          | ||  || |||||  | |::| | :|||:|:|:| ||: | :|:|||||||||||::| 
Db    121 SYRYKLLNWAYQQVQQNKEDAWIEHDVWRMEIYVSLGIVGLAILALLAVTSIPSVSDSLT 180

Qy    388 WREFSFIQSTLGYVALLISTFHVLIYGWKRAFEEEYYRFYTPPNFVLALVLPSIVIL 444
          |||| :||| || |:||: | | ||: | :  : : : :|||| |::|: || :|::
Db    181 WREFHYIQSKLGIVSLLLGTIHALIFAWNKWIDIKQFVWYTPPTFMIAVFLPIVVLI 237


RESULT 9
US-10-012-896-879
; Sequence 879, Application US/10012896
; Publication No. US20020183251A1
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Mitcham, Jennifer L.
;  APPLICANT:  Harlocker, Susan L.
;  APPLICANT:  Jiang, Yuqiu
;  APPLICANT:  Kalos, Michael D.
;  APPLICANT:  Retter, Marc W.
;  APPLICANT:  Stolk, John A.
;  APPLICANT:  Day, Craig H.
;  APPLICANT:  Vedvick, Thomas S.
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Li, Samuel X.
;  APPLICANT:  Wang, Aijun
;  APPLICANT:  Skeiky, Yasir A.W.
;  APPLICANT:  Hepler, William T.
;  APPLICANT:  Henderson, Robert A.
;  APPLICANT:  Hural, John
;  APPLICANT:  McNeill, Patricia D.
;  APPLICANT:  Houghton, Raymond L.
;  APPLICANT:  Vinals de Bassols, Carlota
;  APPLICANT:  Foy, Teresa
;  APPLICANT:  Fanger, Gary R.
;  APPLICANT:  Wantanabe, Yoshihiro
;  APPLICANT:  Meagher, Madeleine Joy
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF PROSTATE CANCER
;  FILE REFERENCE: 210121.427C27
;  CURRENT APPLICATION NUMBER: US/10/012,896
;  CURRENT FILING DATE:  2001-12-10
;  NUMBER OF SEQ ID NOS: 1011
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 879
;   LENGTH: 339
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-012-896-879

  Query Match             30.5%;  Score 717;  DB 9;  Length 339;
  Best Local Similarity   54.9%;  Pred. No. 4.1e-59;
  Matches  130;  Conservative   48;  Mismatches   59;  Indels    0;  Gaps    0;

Qy    208 LFTLWRGPVVVAISLATFFFLYSFVRDVIHPYARNQQSDFYKIPIEIVNKTLPIVAITLL 267
          ||  |  |: :|  :|:  ||| :|:|||| | : |  |||||| ::|| ||:|:|||| 
Db     67 LFPQWHLPIKIAAIIASLTFLYTLLREVIHPLATSHQQYFYKIPILVINKVLPMVSITLL 126

Qy    268 SLVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQCRKQLGLLSFFFAMVHVAYSLCLPMRR 327
          :|||| |::||  ||: ||||::|| ||: |:  ||| |||||||::|  |||  |||| 
Db    127 ALVYLPGVIAAIVQLHNGTKYKKFPHWLDKWMLTRKQFGLLSFFFAVLHAIYSLSYPMRR 186

Qy    328 SERYLFLNMAYQQVHANIENSWNEEEVWRIEMYISFGIMSLGLLSLLAVTSIPSVSNALN 387
          | ||  || |||||  | |::| | :|||:|:|:| ||: | :|:|||||||||||::| 
Db    187 SYRYKLLNWAYQQVQQNKEDAWIEHDVWRMEIYVSLGIVGLAILALLAVTSIPSVSDSLT 246

Qy    388 WREFSFIQSTLGYVALLISTFHVLIYGWKRAFEEEYYRFYTPPNFVLALVLPSIVIL 444
          |||| :||| || |:||: | | ||: | :  : : : :|||| |::|: || :|::
Db    247 WREFHYIQSKLGIVSLLLGTIHALIFAWNKWIDIKQFVWYTPPTFMIAVFLPIVVLI 303


RESULT 10
US-09-802-520-11
; Sequence 11, Application US/09802520
; Publication No. US20020187472A1
; GENERAL INFORMATION:
;  APPLICANT:  Lal, Preeti
;  APPLICANT:  Faris, Mary
;  APPLICANT:  Chen, Huei-Mei
;  APPLICANT:  Ison, Craig H.
;  TITLE OF INVENTION:  STEAP-RELATED PROTEIN
;  FILE REFERENCE:  PC-0037 US
;  CURRENT APPLICATION NUMBER: US/09/802,520
;  CURRENT FILING DATE:  2001-03-09
;  NUMBER OF SEQ ID NOS:  11
;  SOFTWARE:  PERL Program
; SEQ ID NO 11
;   LENGTH: 339
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20020187472A1 g6572948
US-09-802-520-11
```

Query Match 30.5%; Score 717; DB 9; Length 339;
Best Local Similarity 54.9%; Pred. No. 4.1e-59;
Matches 130; Conservative 48; Mismatches 59; Indels 0; Gaps 0;

```
Qy    208 LFTLWRGPVVVAISLATFFFLYSFVRDVIHPYARNQQSDFYKIPIEIVNKTLPIVAITLL 267
          ||  |  |: :|  :|:  |||: :|:|||| | : |  |||||| ::|| ||:|:|||| 
Db     67 LFPQWHLPIKIAAIIASLTFLYTLLREVIHPLATSHQQYFYKIPILVINKVLPMVSITLL 126

Qy    268 SLVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQCRKQLGLLSFFFAMVHVAYSLCLPMRR 327
          :|||| |::||  ||: ||||::|| ||: |:  ||| |||||||::|  |||  |||| 
Db    127 ALVYLPGVIAAIVQLHNGTKYKKFPHWLDKWMLTRKQFGLLSFFFAVLHAIYSLSYPMRR 186

Qy    328 SERYLFLNMAYQQVHANIENSWNEEEVWRIEMYISFGIMSLGLLSLLAVTSIPSVSNALN 387
          | ||  || |||||  | |::| | :|||:|:|:| ||: | :|:|||||||||||::| 
Db    187 SYRYKLLNWAYQQVQQNKEDAWIEHDVWRMEIYVSLGIVGLAILALLAVTSIPSVSDSLT 246

Qy    388 WREFSFIQSTLGYVALLISTFHVLIYGWKRAFEEEYYRFYTPPNFVLALVLPSIVIL 444
          |||| :||| || |:||: | | ||: | :  : : : :|||| |::|: || :|::
Db    247 WREFHYIQSKLGIVSLLLGTIHALIFAWNKWIDIKQFVWYTPPTFMIAVFLPIVVLI 303
```

RESULT 11
US-09-895-793-879
; Sequence 879, Application US/09895793
; Publication No. US20020192763A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.534C2
; CURRENT APPLICATION NUMBER: US/09/895,793
; CURRENT FILING DATE: 2001-06-29
; NUMBER OF SEQ ID NOS: 982
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 879
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-895-793-879

Query Match 30.5%; Score 717; DB 9; Length 339;
Best Local Similarity 54.9%; Pred. No. 4.1e-59;
Matches 130; Conservative 48; Mismatches 59; Indels 0; Gaps 0;

```
Qy    208 LFTLWRGPVVVAISLATFFFLYSFVRDVIHPYARNQQSDFYKIPIEIVNKTLPIVAITLL 267
          ||  |  |: :|  :|:  |||: :|:|||| | : |  |||||| ::|| ||:|:|||| 
Db     67 LFPQWHLPIKIAAIIASLTFLYTLLREVIHPLATSHQQYFYKIPILVINKVLPMVSITLL 126

Qy    268 SLVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQCRKQLGLLSFFFAMVHVAYSLCLPMRR 327
          :|||| |::||  ||: ||||::|| ||: |:  ||| |||||||::|  |||  |||| 
Db    127 ALVYLPGVIAAIVQLHNGTKYKKFPHWLDKWMLTRKQFGLLSFFFAVLHAIYSLSYPMRR 186

Qy    328 SERYLFLNMAYQQVHANIENSWNEEEVWRIEMYISFGIMSLGLLSLLAVTSIPSVSNALN 387
          | ||  || |||||  | |::| | :|||:|:|:| ||: | :|:|||||||||||::| 
Db    187 SYRYKLLNWAYQQVQQNKEDAWIEHDVWRMEIYVSLGIVGLAILALLAVTSIPSVSDSLT 246

Qy    388 WREFSFIQSTLGYVALLISTFHVLIYGWKRAFEEEYYRFYTPPNFVLALVLPSIVIL 444
          |||| :||| || |:||: | | ||: | :  : : : :|||| |::|: || :|::
Db    247 WREFHYIQSKLGIVSLLLGTIHALIFAWNKWIDIKQFVWYTPPTFMIAVFLPIVVLI 303
```

RESULT 12
US-09-895-814-879
; Sequence 879, Application US/09895814
; Publication No. US20020193296A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C26
; CURRENT APPLICATION NUMBER: US/09/895,814
; CURRENT FILING DATE: 2001-06-29
; NUMBER OF SEQ ID NOS: 990
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 879
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-895-814-879

Query Match 30.5%; Score 717; DB 9; Length 339;
Best Local Similarity 54.9%; Pred. No. 4.1e-59;
Matches 130; Conservative 48; Mismatches 59; Indels 0; Gaps 0;

```
Qy    208 LFTLWRGPVVVAISLATFFFLYSFVRDVIHPYARNQQSDFYKIPIEIVNKTLPIVAITLL 267
          ||  |  |: :|  :|:  |||: :|:|||| | : |  |||||| ::|| ||:|:|||| 
Db     67 LFPQWHLPIKIAAIIASLTFLYTLLREVIHPLATSHQQYFYKIPILVINKVLPMVSITLL 126

Qy    268 SLVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQCRKQLGLLSFFFAMVHVAYSLCLPMRR 327
          :|||| |::||  ||: ||||::|| ||: |:  ||| |||||||::|  |||  |||| 
Db    127 ALVYLPGVIAAIVQLHNGTKYKKFPHWLDKWMLTRKQFGLLSFFFAVLHAIYSLSYPMRR 186

Qy    328 SERYLFLNMAYQQVHANIENSWNEEEVWRIEMYISFGIMSLGLLSLLAVTSIPSVSNALN 387
          | ||  || |||||  | |::| | :|||:|:|:| ||: | :|:|||||||||||::| 
Db    187 SYRYKLLNWAYQQVQQNKEDAWIEHDVWRMEIYVSLGIVGLAILALLAVTSIPSVSDSLT 246

Qy    388 WREFSFIQSTLGYVALLISTFHVLIYGWKRAFEEEYYRFYTPPNFVLALVLPSIVIL 444
          |||| :||| || |:||: | | ||: | :  : : : :|||| |::|: || :|::
Db    247 WREFHYIQSKLGIVSLLLGTIHALIFAWNKWIDIKQFVWYTPPTFMIAVFLPIVVLI 303
```

RESULT 13
US-09-759-143-879
; Sequence 879, Application US/09759143

```
; Patent No. US20020022248A1
; GENERAL INFORMATION:
;  APPLICANT:  Xu, Jiangchun
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Mitcham, Jennifer L.
;  APPLICANT:  Harlocker, Susan L.
;  APPLICANT:  Jiang, Yuqui
;  APPLICANT:  Henderson, Robert A.
;  APPLICANT:  Kalos, Michael D.
;  APPLICANT:  Fanger, Gary R.
;  APPLICANT:  Retter, Marc W.
;  APPLICANT:  Stolk, John A.
;  APPLICANT:  Day, Craig H.
;  APPLICANT:  Vedvick, Thomas S.
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Li, Samuel
;  APPLICANT:  Wang, Aijun
;  APPLICANT:  Skeiky, Yasir A.W.
;  APPLICANT:  Hepler, William
;  TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF PROSTATE CANCER
;  FILE REFERENCE:  210121.427C23
;  CURRENT APPLICATION NUMBER: US/09/759,143
;  CURRENT FILING DATE:  2001-01-12
;  NUMBER OF SEQ ID NOS:  934
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 879
;   LENGTH: 339
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-759-143-879

  Query Match             30.5%;  Score 717;  DB 10;  Length 339;
  Best Local Similarity   54.9%;  Pred. No. 4.1e-59;
  Matches  130;  Conservative   48;  Mismatches   59;  Indels    0;  Gaps    0;

Qy     208 LFTLWRGPVVVAISLATFFFLYSFVRDVIHPYARNQQSDFYKIPIEIVNKTLPIVAITLL 267
           ||  |  |: :|  :|:  |||: :|:|||| | : |  |||||| ::|| ||:|:|||| 
Db      67 LFPQWHLPIKIAAIIASLTFLYTLLREVIHPLATSHQQYFYKIPILVINKVLPMVSITLL 126

Qy     268 SLVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQCRKQLGLLSFFFAMVHVAYSLCLPMRR 327
           :|||| |::||  ||: ||||::|| ||: |:  ||| |||||||::|  |||  |||| 
Db     127 ALVYLPGVIAAIVQLHNGTKYKKFPHWLDKWMLTRKQFGLLSFFFAVLHAIYSLSYPMRR 186

Qy     328 SERYLFLNMAYQQVHANIENSWNEEEVWRIEMYISFGIMSLGLLSLLAVTSIPSVSNALN 387
           | ||  || |||||  | |::| | :|||:|:|:| ||: | :|:|||||||||||::|
Db     187 SYRYKLLNWAYQQVQQNKEDAWIEHDVWRMEIYVSLGIVGLAILALLAVTSIPSVSDSLT 246

Qy     388 WREFSFIQSTLGYVALLISTFHVLIYGWKRAFEEEYYRFYTPPNFVLALVLPSIVIL 444
           |||| :||| || |:||: | | ||: | :  : : : :|||| |::|: || :|::
Db     247 WREFHYIQSKLGIVSLLLGTIHALIFAWNKWIDIKQFVWYTPPTFMIAVFLPIVVLI 303


RESULT 14
US-09-780-669-879
; Sequence 879, Application US/09780669
; Patent No. US20020051977A1
; GENERAL INFORMATION:
;  APPLICANT:  Xu, Jiangchun
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Mitcham, Jennifer L.
;  APPLICANT:  Harlocker, Susan L.
;  APPLICANT:  Jiang, Yuqui
;  APPLICANT:  Henderson, Robert A.
;  APPLICANT:  Kalos, Michael D.
;  APPLICANT:  Fanger, Gary R.
;  APPLICANT:  Retter, Marc W.
;  APPLICANT:  Stolk, John A.
;  APPLICANT:  Day, Craig H.
;  APPLICANT:  Vedvick, Thomas S.
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Li, Samuel
;  APPLICANT:  Wang, Aijun
;  APPLICANT:  Skeiky, Yasir A.W.
;  APPLICANT:  Hepler, William
;  APPLICANT:  Hural, John
;  APPLICANT:  McNeill, Patricia D.
;  APPLICANT:  Houghton, Raymond L.
;  TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF PROSTATE CANCER
;  FILE REFERENCE:  210121.427C24
;  CURRENT APPLICATION NUMBER: US/09/780,669
;  CURRENT FILING DATE:  2001-02-09
;  NUMBER OF SEQ ID NOS:  943
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 879
;   LENGTH: 339
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-780-669-879

  Query Match             30.5%;  Score 717;  DB 10;  Length 339;
  Best Local Similarity   54.9%;  Pred. No. 4.1e-59;
  Matches  130;  Conservative   48;  Mismatches   59;  Indels    0;  Gaps    0;

Qy     208 LFTLWRGPVVVAISLATFFFLYSFVRDVIHPYARNQQSDFYKIPIEIVNKTLPIVAITLL 267
           ||  |  |: :|  :|:  |||: :|:|||| | : |  |||||| ::|| ||:|:|||| 
Db      67 LFPQWHLPIKIAAIIASLTFLYTLLREVIHPLATSHQQYFYKIPILVINKVLPMVSITLL 126

Qy     268 SLVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQCRKQLGLLSFFFAMVHVAYSLCLPMRR 327
           :|||| |::||  ||: ||||::|| ||: |:  ||| |||||||::|  |||  |||| 
Db     127 ALVYLPGVIAAIVQLHNGTKYKKFPHWLDKWMLTRKQFGLLSFFFAVLHAIYSLSYPMRR 186

Qy     328 SERYLFLNMAYQQVHANIENSWNEEEVWRIEMYISFGIMSLGLLSLLAVTSIPSVSNALN 387
           | ||  || |||||  | |::| | :|||:|:|:| ||: | :|:|||||||||||::|
Db     187 SYRYKLLNWAYQQVQQNKEDAWIEHDVWRMEIYVSLGIVGLAILALLAVTSIPSVSDSLT 246

Qy     388 WREFSFIQSTLGYVALLISTFHVLIYGWKRAFEEEYYRFYTPPNFVLALVLPSIVIL 444
           |||| :||| || |:||: | | ||: | :  : : : :|||| |::|: || :|::
Db     247 WREFHYIQSKLGIVSLLLGTIHALIFAWNKWIDIKQFVWYTPPTFMIAVFLPIVVLI 303


RESULT 15
US-09-822-827-879
; Sequence 879, Application US/09822827
; Patent No. US20020081680A1
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF PROSTATE CANCER
;  FILE REFERENCE: 210121.534C1
;  CURRENT APPLICATION NUMBER: US/09/822,827
;  CURRENT FILING DATE:  2001-03-28
;  NUMBER OF SEQ ID NOS: 982
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 879
;   LENGTH: 339
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-822-827-879

  Query Match             30.5%;  Score 717;  DB 10;  Length 339;
  Best Local Similarity   54.9%;  Pred. No. 4.1e-59;
  Matches  130;  Conservative   48;  Mismatches   59;  Indels    0;  Gaps    0;

Qy     208 LFTLWRGPVVVAISLATFFFLYSFVRDVIHPYARNQQSDFYKIPIEIVNKTLPIVAITLL 267
           ||  |  |: :|  :|:  |||: :|:|||| | : |  |||||| ::|| ||:|:|||| 
Db      67 LFPQWHLPIKIAAIIASLTFLYTLLREVIHPLATSHQQYFYKIPILVINKVLPMVSITLL 126

Qy     268 SLVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQCRKQLGLLSFFFAMVHVAYSLCLPMRR 327
           :|||| |::||  ||: ||||::|| ||: |:  ||| |||||||::|  |||  |||| 
Db     127 ALVYLPGVIAAIVQLHNGTKYKKFPHWLDKWMLTRKQFGLLSFFFAVLHAIYSLSYPMRR 186

Qy     328 SERYLFLNMAYQQVHANIENSWNEEEVWRIEMYISFGIMSLGLLSLLAVTSIPSVSNALN 387
           | ||  || |||||  | |::| | :|||:|:|:| ||: | :|:|||||||||||::|
Db     187 SYRYKLLNWAYQQVQQNKEDAWIEHDVWRMEIYVSLGIVGLAILALLAVTSIPSVSDSLT 246
```

```
Qy     388 WREFSFIQSTLGYVALLISTFHVLIYGWKRAFEEEYYRFYTPPNFVLALVLPSIVIL 444
           |||| :||| || |:||: | | ||: | :  : : : :|||| |::|: || :|::
Db     247 WREFHYIQSKLGIVSLLLGTIHALIFAWNKWIDIKQFVWYTPPTFMIAVFLPIVVLI 303
```

Search completed: January 17, 2003, 19:04:43
Job time : 18 secs

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: January 17, 2003, 18:58:14 ; Search time 14 Seconds
(without alignments)
1345.018 Million cell updates/sec

Title: US-09-455-486-6
Perfect score: 2351
Sequence: 1 MESISMMGSPKSLSETCLPN.....ALVLPSIVILDLLQLCRYPD 454

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	717	30.5	339	1	STEA_HUMAN	Q9uhe8 homo sapien
2	133	5.7	223	1	F4RE_METJA	Q58896 methanococc
3	120.5	5.1	232	1	F4RE_METTH	O26350 methanobact
4	111.5	4.7	695	1	FSHR_MACFA	P32212 macaca fasc
5	109.5	4.7	206	1	Y538_PASMU	Q9cn97 pasteurella
6	109	4.6	693	1	FSHR_CHICK	P79763 gallus gall
7	109	4.6	712	1	FRE6_YEAST	Q12473 saccharomyc
8	107.5	4.6	629	1	FRE7_YEAST	Q12333 saccharomyc
9	107.5	4.6	695	1	FSHR_HUMAN	P23945 homo sapien
10	104.5	4.4	695	1	FSHR_PIG	P49059 sus scrofa
11	104	4.4	396	1	SOTB_ECO57	P58529 escherichia
12	104	4.4	396	1	SOTB_ECOLI	P31122 escherichia
13	103	4.4	465	1	NPT1_MOUSE	Q61983 mus musculu
14	102	4.3	220	1	Y304_BRUME	Q8yd73 brucella me
15	100	4.3	476	1	RF3_SACBA	P05512 saccharomyc
16	99.5	4.2	452	1	NU4M_BRALA	O79421 branchiosto
17	99	4.2	365	1	GP68_HUMAN	Q15743 homo sapien
18	98	4.2	3411	1	POLG_YEFV1	P03314 y genome po
19	98	4.2	3411	1	POLG_YEFV2	P19901 y genome po
20	97.5	4.1	268	1	YD49_AQUAE	O67364 aquifex aeo
21	97.5	4.1	695	1	FSHR_BOVIN	P35376 bos taurus
22	97	4.1	692	1	FSHR_RAT	P20395 rattus norv
23	96.5	4.1	686	1	FRE1_YEAST	P32791 saccharomyc
24	96	4.1	261	1	PHSC_ECOLI	P77409 escherichia
25	95	4.0	202	1	YAJ1_PSEAE	Q9hva5 pseudomonas
26	95	4.0	393	1	NUCC_ARATH	P56753 arabidopsis
27	95	4.0	393	1	NUCC_SPIOL	Q9m3i5 spinacia ol
28	95	4.0	692	1	FSHR_MOUSE	P35378 mus musculu
29	94.5	4.0	492	1	SECY_CYAPA	P25014 cyanophora
30	94.5	4.0	695	1	FSHR_SHEEP	P35379 ovis aries
31	93.5	4.0	452	1	NU4M_BRAFL	O47423 branchiosto
32	93.5	4.0	499	1	GSHR_PLAF7	O15770 plasmodium
33	93.5	4.0	1233	1	VL1_REOVD	P15024 reovirus (t
34	93	4.0	388	1	HMC3_DESVH	P33390 desulfovibr
35	93	4.0	459	1	NU4M_BOVIN	P03910 bos taurus
36	93	4.0	521	1	YT25_CAEEL	Q10934 caenorhabdi
37	92.5	3.9	490	1	GTR1_CHICK	P46896 gallus gall
38	92.5	3.9	500	1	NU4C_OENHO	P58419 oenothera h
39	92.5	3.9	883	1	YHL6_YEAST	P38781 saccharomyc
40	92.5	3.9	1769	1	YJK9_YEAST	P42945 saccharomyc
41	92.5	3.9	2037	1	FAS1_CANAL	P34731 c fatty aci
42	91.5	3.9	495	1	TPS1_YEAST	Q00764 s alpha,alp
43	91.5	3.9	1694	1	CLH_DICDI	P25870 dictyosteli
44	91	3.9	241	1	Y705_CAMJE	Q9ppk2 campylobact
45	91	3.9	409	1	HEM1_THEAC	Q9hkr0 thermoplasm

ALIGNMENTS

RESULT 1
STEA_HUMAN
ID STEA_HUMAN STANDARD; PRT; 339 AA.
AC Q9UHE8; O95034;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Six transmembrane epithelial antigen of prostate.
GN STEAP OR STEAP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20056277; PubMed=10588738;
RA Hubert R.S., Vivanco I., Chen E., Rastegar S., Leong K.,
RA Mitchell S.C., Madraswala R., Zhou Y., Kuo J., Raitano A.B.,
RA Jakobovits A., Saffran D.C., Afar D.E.H.;
RT "STEAP: a prostate-specific cell-surface antigen highly expressed in
RT human prostate tumors.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:14523-14528(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Abu-Threideh J., Stoneking T., Langston Y., Maupin R.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Skin;
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN PROSTATE TUMORS.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; AF186249; AAF17479.1; -.
DR EMBL; AC005053; AAC79150.1; ALT_INIT.
DR EMBL; AC004969; AAD15620.1; ALT_INIT.
DR EMBL; BC011802; AAH11802.1; -.
DR Genew; HGNC:11378; STEAP.
DR MIM; 604415; -.
KW Transmembrane; Antigen.
FT TRANSMEM 71 91 POTENTIAL.
FT TRANSMEM 119 139 POTENTIAL.
FT TRANSMEM 164 184 POTENTIAL.
FT TRANSMEM 218 238 POTENTIAL.
FT TRANSMEM 258 278 POTENTIAL.
FT TRANSMEM 291 311 POTENTIAL.
SQ SEQUENCE 339 AA; 39851 MW; 55443A170C870387 CRC64;

```
  Query Match             30.5%;  Score 717;  DB 1;  Length 339;
  Best Local Similarity   54.9%;  Pred. No. 7e-44;
  Matches  130;  Conservative   48;  Mismatches   59;  Indels    0;  Gaps    0;

Qy     208 LFTLWRGPVVVAISLATFFFLYSFVRDVIHPYARNQQSDFYKIPIEIVNKTLPIVAITLL 267
           ||  |  |: :|  :|:  ||| :|:|||| | : |  |||||| ::|| ||:|:|||| 
Db      67 LFPQWHLPIKIAAIIASLTFLYTLLREVIHPLATSHQQYFYKIPILVINKVLPMVSITLL 126

Qy     268 SLVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQCRKQLGLLSFFFAMVHVAYSLCLPMRR 327
           :|||| |::|| ||: |||||::|| ||: |:  ||| |||||||||::|  |||  |||| 
Db     127 ALVYLPGVIAAIVQLHNGTKYKKFPHWLDKWMLTRKQFGLLSFFFAVLHAIYSLSYPMRR 186

Qy     328 SERYLFLNMAYQQVHANIENSWNEEEVWRIEMYISFGIMSLGLLSLLAVTSIPSVSNALN 387
           | ||  || |||||  | |::| | :|||:|:|:| ||: | :|:|||||||||||::| 
Db     187 SYRYKLLNWAYQQVQQNKEDAWIEHDVWRMEIYVSLGIVGLAILALLAVTSIPSVSDSLT 246

Qy     388 WREFSFIQSTLGYVALLISTFHVLIYGWKRAFEEEYYRFYTPPNFVLALVLPSIVIL 444
           |||| :||| || |:||: | | ||: | :  : : : :|||| |::|: || :|::
Db     247 WREFHYIQSKLGIVSLLLGTIHALIFAWNKWIDIKQFVWYTPPTFMIAVFLPIVVLI 303


RESULT 2
F4RE_METJA
ID   F4RE_METJA     STANDARD;      PRT;   223 AA.
AC   Q58896;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Putative F420-dependent NADP reductase (EC 1.-.-.-).
GN   MJ1501.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC   Methanocaldococcaceae; Methanocaldococcus.
OX   NCBI_TaxID=2190;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX   MEDLINE=96337999; PubMed=8688087;
RA   Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA   Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA   Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA   Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA   Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA   Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA   Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA   Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT   jannaschii.";
RL   Science 273:1058-1073(1996).
CC   -!- FUNCTION: CATALYSES THE REDUCTION OF F420 WITH NADP(+) AND THE
CC       REDUCTION OF NADP(+) WITH F420H(2). PROBABLY COUPLES THE NADP-
CC       DEPENDENT OXIDATION OF THE ALCOHOL TO THE ALDEHYDE WITH DE F420-
CC       DEPENDENT REDUCTION OF CO(2) TO METHANE (ANABOLIC FUNCTION) (BY
CC       SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: NADPH + coenzyme F420 = NADP(+) + reduced
CC       coenzyme F420.
CC   -!- SIMILARITY: TO M.ORGANOPHILUM F420-DEPENDENT NADP REDUCTASE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U67591; AAB99514.1; -.
DR   TIGR; MJ1501; -.
DR   InterPro; IPR004455; NADPoxred_F420.
DR   TIGRFAMs; TIGR00301; NADPoxidored_F420; 1.
KW   Hypothetical protein; Oxidoreductase; NADP; Complete proteome.
SQ   SEQUENCE   223 AA;  24068 MW;  2370BBD6F5BBD3D9 CRC64;

  Query Match             5.7%;  Score 133;  DB 1;  Length 223;
  Best Local Similarity   21.9%;  Pred. No. 0.01;
  Matches   46;  Conservative   52;  Mismatches   78;  Indels   34;  Gaps    7;

Qy      37 GSGDFAKSLTIRLIRCGYHVVIGSRNPKFASEFFPHVVDV----------THHEDALT 84
           |:||    | :|| :    ::|||| : | |      ::             ::||
Db       8 GTGDQGFGLALRLAK-NNKIIIGSRKKEKAEEAAKKAKEILKQRGIEADIIGLENKDAAK 66

Qy      85 KTNIIFVAIHREH-YTSLWDLRHLLVGKILIDV---------SNNMRINQYPE-SNAEYL 133
           : ::: ::: |: ::: |: | |||:: :             |:   |: | || :
Db      67 EGDVVILSLPYEYTLSTIKQLKEELKGKIVVSIGVPLATAIGDKPTRLLFPPDGSVAEMV 126

Qy     134 ASLFPDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIELARQLNFI-PIDL 192
            ::  :| :|  |  |    |:         : :| |: :|:: ||:|| |:: :  ||
Db     127 QNVLKESKVVSAFQNVCHAVLEDLDNPVDCDILVCGNDEEAKKVVIDLANQIDGVRAIDC 186

Qy     193 GSLSSAREIENLPLRLFTLWRGPVVVAISL 222
           |:|  :| || :          |::: :::
Db     187 GNLEKSRIIEAIT---------PLLIGLNI 207


RESULT 3
F4RE_METTH
ID   F4RE_METTH     STANDARD;      PRT;   232 AA.
AC   O26350;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Putative F420-dependent NADP reductase (EC 1.-.-.-).
GN   MTH248.
OS   Methanobacterium thermoautotrophicum.
OC   Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC   Methanobacteriaceae; Methanothermobacter.
OX   NCBI_TaxID=187420;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Delta H;
RX   MEDLINE=98037514; PubMed=9371463;
RA   Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA   Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA   Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA   Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA   Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA   McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
RA   Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT   "Complete genome sequence of Methanobacterium thermoautotrophicum
RT   deltaH: functional analysis and comparative genomics.";
RL   J. Bacteriol. 179:7135-7155(1997).
CC   -!- FUNCTION: CATALYSES THE REDUCTION OF F420 WITH NADP(+) AND THE
CC       REDUCTION OF NADP(+) WITH F420H(2). PROBABLY COUPLES THE NADP-
CC       DEPENDENT OXIDATION OF THE ALCOHOL TO THE ALDEHYDE WITH DE F420-
CC       DEPENDENT REDUCTION OF CO(2) TO METHANE (ANABOLIC FUNCTION) (BY
CC       SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: NADPH + coenzyme F420 = NADP(+) + reduced
CC       coenzyme F420.
CC   -!- SIMILARITY: TO M.ORGANOPHILUM F420-DEPENDENT NADP REDUCTASE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000811; AAB84754.1; -.
DR   InterPro; IPR004455; NADPoxred_F420.
DR   TIGRFAMs; TIGR00301; NADPoxidored_F420; 1.
KW   Hypothetical protein; Oxidoreductase; NADP; Complete proteome.
SQ   SEQUENCE   232 AA;  24539 MW;  A1CE60ABC8474296 CRC64;

  Query Match             5.1%;  Score 120.5;  DB 1;  Length 232;
```

```
  Best Local Similarity   23.8%;  Pred. No. 0.084;
  Matches   49;  Conservative   42;  Mismatches   90;  Indels   25;  Gaps    6;

Qy      30 KVTVGVIGSGDFAKSLTIRLIRCGYHVVIGSRNPKFASEFFPHVVDVTHHE--------- 80
           |: | : |:||    | :|    |  |:||||: : ||:    |:::   :
Db       8 KIAV-IGGTGDQGLGLALRFAVAGEEVIIGSRDAEKASKAASKVLEIAGRDDISVEGATN 66

Qy      81 -DALTKTNIIFVAIH-REHYTSLWDLRHLLVGKILIDVS---------NNMRINQYPESN 129
            ||    ::: : :  :    :|  :|  :  |:||| :         : :|     | :
Db      67 PDAAASADVVVLTVPLQAQMVTLASIRDQVRDKVLIDATVPIDSCIGGSAVRYIDLWEGS 126

Qy     130 AEYLASLF---PDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIELARQLN 186
           |   |: |     : :   || :|| ||     :       : |::  | :   ||| :::
Db     127 AAERAARFLREQGTRVAAAFNNISASALLEVSEPVDCDCLVASDHRDALEVAAELAEKID 186

Qy     187 FI-PIDLGSLSSAREIENLPLRLFTL 211
           :  |: | | :|| || :   |  |
Db     187 GVRAIECGGLENARIIEKITPLLINL 212


RESULT 4
FSHR_MACFA
ID   FSHR_MACFA     STANDARD;      PRT;   695 AA.
AC   P32212;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Follicle stimulating hormone receptor precursor (FSH-R) (Follitropin
DE   receptor).
GN   FSHR.
OS   Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9541;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RX   MEDLINE=94071854; PubMed=7504463;
RA   Gromoll J., Dankbar B., Sharma R.S., Nieschlag E.;
RT   "Molecular cloning of the testicular follicle stimulating hormone
RT   receptor of the non human primate Macaca fascicularis and
RT   identification of multiple transcripts in the testis.";
RL   Biochem. Biophys. Res. Commun. 196:1066-1072(1993).
CC   -!- FUNCTION: RECEPTOR FOR FOLLICLE STIMULATING HORMONE. THE ACTIVITY
CC       OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE
CC       ADENYLATE CYCLASE.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC       FSH/LSH/TSH SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 6 LEUCINE-RICH REPEATS (LRR).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X74454; CAA52463.1; -.
DR   PIR; S36452; S36452.
DR   PIR; JN0898; JN0898.
DR   HSSP; P23945; 1XUN.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000372; LRR_Nterm.
DR   Pfam; PF00001; 7tm_1; 1.
DR   Pfam; PF00560; LRR; 3.
DR   Pfam; PF01462; LRRNT; 1.
DR   SMART; SM00013; LRRNT; 1.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
KW   Phosphorylation; Repeat; Leucine-rich repeat.
FT   SIGNAL       1     17       POTENTIAL.
FT   CHAIN       18    695       FOLLICLE STIMULATING HORMONE RECEPTOR.
FT   DOMAIN      18    366       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   367    387       1 (POTENTIAL).
FT   DOMAIN     388    398       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   399    421       2 (POTENTIAL).
FT   DOMAIN     422    443       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   444    465       3 (POTENTIAL).
FT   DOMAIN     466    485       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   486    508       4 (POTENTIAL).
FT   DOMAIN     509    528       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   529    550       5 (POTENTIAL).
FT   DOMAIN     551    573       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   574    597       6 (POTENTIAL).
FT   DOMAIN     598    608       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   609    630       7 (POTENTIAL).
FT   DOMAIN     631    695       CYTOPLASMIC (POTENTIAL).
FT   REPEAT      44     68       LRR 1.
FT   REPEAT      69     93       LRR 2.
FT   REPEAT     119    143       LRR 3.
FT   REPEAT     170    192       LRR 4.
FT   REPEAT     193    216       LRR 5.
FT   REPEAT     218    240       LRR 6.
FT   DISULFID   442    517       BY SIMILARITY.
FT   CARBOHYD   191    191       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   199    199       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   293    293       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   318    318       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   695 AA;  78343 MW;  0D60A233729B5250 CRC64;

  Query Match             4.7%;  Score 111.5;  DB 1;  Length 695;
  Best Local Similarity   17.0%;  Pred. No. 1.3;
  Matches   84;  Conservative   75;  Mismatches  151;  Indels  183;  Gaps   16;

Qy      79 HEDALTKTNIIFVAIHREHYTSLWDLRHLLVG----------------KILIDVSNNMR 121
           ||  : | | : : |: | : :| :||:||:                 |:|:|: :|:
Db      98 HEIRIEKANNL-LYINPEAFQNLPNLRYLLISNTGIKHLPDVHKIHSFQKVLLDIQDNIN 156

Qy     122 INQYPESNAEYLASLFPDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIEL 181
           |:          ::  :| :   |  |  |     |:  :::: |:
Db     157 IH-----------TIERNSFVGLSFESVILWL----NKNGIQEIHNCA---------- 189

Qy     182 ARQLNFIPIDLGSLSSAREIENLPLRLFTLWRGPVVVAISLATFFFLYSFVRDVIHPYAR 241
               |   :|  :||    :| ||  :|    |||:: ||      | |:  : :
Db     190 ---FNGTQLDELNLSDNNNLEELPNDVFHGASGPVILDISRTRIHSLPSYGLENLKKLRA 246

Qy     242 NQQSDFYKIPIEIVNKTLPIVAITLLSLVY---------------------------- 271
           :  |:|         :||:   || |
Db     247 RSTYNLKKLP-----SLEKLVALMEASLTYPSHCCAFANWRRQISELHPICNKSILRQEV 301

Qy     272 ------------LAGLLAAAYQLYYGTKYRRF------------PPWLETWLQCRKQLG 306
                    ||    ::|   :   |  |              | : : |   ||
Db     302 DYMTQTRGQRSSLAEDNESSYSRGFDMTYAEFDYDLCNEVVDVTCSPKPDAFNPCEDILG 361

Qy     307 -----LLSFFFAMVHVA------------YSLCLPMRRSERYLFLNMAYQQVHANIENS 348
               :| :| ::: :              | | :|     |:|  |:|:
Db     362 YNILRVLIWFISILAITGNIIVLVTLTTSQYKLTVP-----RFLMCNLAFAD-------- 408

Qy     349 WNEEEVWRIEMYISFGIMSLGLLSLLAVTSIPSVSNALNWRE--------FSFIQSTLG 399
                  :  ||  | : |:    |     : |::|:          |:   | |
Db     409 ------------LCIGIYLLLIASVDIHTKSQYHNYAIDWQTGAGCDAAGFFTVFASELS 456

Qy     400 ---YVALLISTFH------------------VLIYGWKRAFEEEYYRFYTPPNFV-LAL 436
             |: :  :|                   |:: ||  ||    :  :   ::: :::
Db     457 VYTLTAITLERWHTITHAMQLDCKVHVRHAASVMVMGWIFAFAAALFPIFGISSYMKVSI 516

Qy     437 VLPSIVILDLLQL 449
           ||  :   | ||
Db     517 CLPMDIDSPLSQL 529
```

```
RESULT 5
Y538_PASMU
ID   Y538_PASMU     STANDARD;      PRT;   206 AA.
AC   Q9CN97;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical protein PM0538.
GN   PM0538.
OS   Pasteurella multocida.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Pasteurella.
OX   NCBI_TaxID=747;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Pm70;
RX   MEDLINE=21145866; PubMed=11248100;
RA   May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT   "Complete genomic sequence of Pasteurella multocida Pm70.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: BELONGS TO THE UPF0191 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE006089; AAK02622.1; -.
KW   Hypothetical protein; Transmembrane; Complete proteome.
FT   TRANSMEM      7     26       POTENTIAL.
FT   TRANSMEM     41     63       POTENTIAL.
FT   TRANSMEM     76     98       POTENTIAL.
FT   TRANSMEM    113    135       POTENTIAL.
FT   TRANSMEM    142    164       POTENTIAL.
FT   TRANSMEM    169    191       POTENTIAL.
SQ   SEQUENCE   206 AA;  23653 MW;  2F277729D7F11194 CRC64;

  Query Match             4.7%;  Score 109.5;  DB 1;  Length 206;
  Best Local Similarity   23.8%;  Pred. No. 0.44;
  Matches   56;  Conservative   44;  Mismatches   72;  Indels   63;  Gaps   13;

Qy     228 LYSFVRDVIH-----PYA-------RNQQSDFYKIPIEIVNKTLPIVAITLLSLVYLAGL 275
           : |  | :||     | |          :|     ||: :   |   |:|:| :::: |
Db       1 MLSLFRIIIHVCCLGPVAWLAWVLLSGDESQLGADPIKEIQHFLGFSALTILLIMFILG- 59

Qy     276 LAAAYQLYYGTKYRRFPPWLETWLQCRKQLGLLSFFFAMVHVAYSLCLPMRRSERYLFLN 335
           :::|  |     | |:     |: ||| ::|: ::|| |:           || |
Db      60 -----KVFYLLK----QPQLQV---LRRALGLWAWFYVVLHV-YA----------YLALE 96

Qy     336 MAYQQVHANIENSWNEEEVWRIEMYISFGIMSLGLLSLLAVTSIPSVSNALNWREFSFIQ 395
           : |        : : :| | |   |:  | ::  :|:|:|::|   :   |   :: |
Db      97 LGYD------FSLFVQELVNR--GYLIIGAIAFLILTLMALSSWSYLK--LKMGKWWFYL 146

Qy     396 STLGYVALLISTFHVLIYGWKRAFEEEYYRFYTPPNFVLALVLPSIVILDLLQLC 450
             ||| |||:   |   | |            :  |   : :|   :|| :: ||
Db     147 HQLGYYALLLGAIH---YVW-----------SVKNVTFSSML--YLILSIMILC 184


RESULT 6
FSHR_CHICK
ID   FSHR_CHICK     STANDARD;      PRT;   693 AA.
AC   P79763; Q90719;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Follicle stimulating hormone receptor precursor (FSH-R) (Follitropin
DE   receptor).
GN   FSHR.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Ovary;
RX   MEDLINE=97473503; PubMed=9332357;
RA   Wakabayashi N., Suzuki A., Hoshino H., Nishimori K., Mizuno S.;
RT   "The cDNA cloning and transient expression of a chicken gene encoding
RT   a follicle-stimulating hormone receptor.";
RL   Gene 197:121-127(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=White leghorn;
RX   MEDLINE=97057887; PubMed=8902217;
RA   You S., Bridgham J.T., Foster D.N., Johnson A.L.;
RT   "Characterization of the chicken follicle-stimulating hormone
RT   receptor (cFSH-R) complementary deoxyribonucleic acid, and expression
RT   of cFSH-R messenger ribonucleic acid in the ovary.";
RL   Biol. Reprod. 55:1055-1062(1996).
CC   -!- FUNCTION: RECEPTOR FOR FOLLICLE STIMULATING HORMONE. THE ACTIVITY
CC       OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE
CC       ADENYLATE CYCLASE.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC       FSH/LSH/TSH SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 7 LEUCINE-RICH REPEATS (LRR).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D87871; BAA13487.1; -.
DR   EMBL; U51097; AAC60030.1; -.
DR   HSSP; P23945; 1XUN.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000372; LRR_Nterm.
DR   Pfam; PF00001; 7tm_1; 1.
DR   Pfam; PF00560; LRR; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   SMART; SM00013; LRRNT; 1.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
KW   Phosphorylation; Repeat; Leucine-rich repeat.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18    693       FOLLICLE STIMULATING HORMONE RECEPTOR.
FT   DOMAIN       18    366       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    367    387       1 (POTENTIAL).
FT   DOMAIN      388    398       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    399    421       2 (POTENTIAL).
FT   DOMAIN      422    443       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    444    465       3 (POTENTIAL).
FT   DOMAIN      466    485       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    486    508       4 (POTENTIAL).
FT   DOMAIN      509    528       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    529    550       5 (POTENTIAL).
FT   DOMAIN      551    573       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    574    597       6 (POTENTIAL).
FT   DOMAIN      598    608       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    609    630       7 (POTENTIAL).
FT   DOMAIN      631    693       CYTOPLASMIC (POTENTIAL).
FT   REPEAT       45     68       LRR 1.
FT   REPEAT       69     93       LRR 2.
FT   REPEAT       95    118       LRR 3.
FT   REPEAT      119    143       LRR 4.
FT   REPEAT      168    192       LRR 5.
```

```
FT   REPEAT      193    216       LRR 6.
FT   REPEAT      218    240       LRR 7.
FT   DISULFID    442    517       BY SIMILARITY.
FT   CARBOHYD     47     47       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    191    191       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    199    199       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    268    268       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT      4      4       G -> D (IN REF. 2).
FT   CONFLICT     88     88       G -> A (IN REF. 2).
FT   CONFLICT    140    140       K -> R (IN REF. 2).
FT   CONFLICT    174    174       I -> T (IN REF. 2).
FT   CONFLICT    191    191       N -> S (IN REF. 2).
FT   CONFLICT    329    329       V -> L (IN REF. 2).
SQ   SEQUENCE   693 AA;  78697 MW;  46F98699635A1BEC CRC64;

  Query Match             4.6%;  Score 109;  DB 1;  Length 693;
  Best Local Similarity   17.2%;  Pred. No. 2;
  Matches  100;  Conservative   72;  Mismatches  173;  Indels  238;  Gaps   21;

Qy      18 LPNG-INGIKDARKVTVG-----VIGSGDFAKSLTIRLIRCGYHVVIGSRNPKFASEFFP 71
           :| |   |: |  |: :       |  |:   ||                 ||
Db      61 IPKGAFTGLHDLEKIEISQNDALEIIEGNVFSSL----------------PKL------ 97

Qy      72 HVVDVTHHEDALTKTNIIFVAIHREHYTSLWDLRHLLVG----------------KILI 114
              ||  : | | : : | :: :  |  ||:||:                    |:|:
Db      98 -------HEIRIEKANKL-MKIDQDAFQHLPSLRYLLISNTGLSFLPVVHKVHSFQKVLL 149

Qy     115 DVSNNMRINQYPESNAEYLASLFPDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQA 174
           || :|: |      :    |:|   :|:|::                    :  | ||
Db     150 DVQDNIHIRTIERNTFMGLSS---ESVILR---------------------LNKNGIQE 184

Qy     175 RQQVIELARQLNFIPIDLGSLSSAREIENLPLRLFTLWRGPVVVAISLATFFFLYSFVRD 234
             :        |   :|  :||    :| || ::|    ||||: ||     || |   :
Db     185 IKD-----HAFNGTCLDELNLSDNYNLEKLPEKVFQGAIGPVVLDISRTRISFLPSHGLE 239

Qy     235 VIHPYARNQQSDFYKIPIEIVNKTLPIVAITLLSLVYLAGL------------------- 275
           |          |:|   |||   ::       :
Db     240 FIKKLRARSTYKLKKLP--DVNKFRSLIEANFTYPSHCCAFTNRKTQNTEFYPICSMSPA 297

Qy     276 ---------------LAAAYQLYYGTKY----RRF-------------PPWLETWLQCR 302
                           |  |  :|||::      |             |  : :  |
Db     298 KQDLGEQTGKRKHRRSAAEDYISHYGTRFGPVENEFDYGLCNEVVDFVCSPKPDAFNPCE 357

Qy     303 KQLG-----LLSFFFAMVHVA-------------YSLCLPMRRSERYLFLNMAYQQVHAN 344
             :|     :| :|  :: :              | | :|     |:|  |:|:
Db     358 DIMGYNVLRVLIWFINILAITGNTTVLIILISSQYKLTVP-----RFLMCNLAFAD---- 408

Qy     345 IENSWNEEEVWRIEMYISFGIMSLGLLSLLAVTSIPSVSNALNWRE---------FSFIQ 395
                      :  ||  | : |:   |     : |::|:          |:
Db     409 ---------------LCIGIYLLFIASVDIQTKSRYYNYAIDWQTGAGCNAAGFFTVFA 452

Qy     396 STLGYVALLISTFH----------------------VLIYGWKRAF-------------- 419
           | |    | : |                        ::::||  ||
Db     453 SELSVYTLTVITLERWHTITYAMQLNRKVRLRHAVIIMVFGWMFAFTVALLPIFGISSYM 512

Qy     420 ----------EEEYYRFYTPPNFVLALVLPSIVILDLLQLCRY 452
                    |  : : |    |:| | : : ||: :  :| |
Db     513 KVSICLPMHIETPFSQAYV--IFLLVLNVLAFVIICICYICIY 553


RESULT 7
FRE6_YEAST
ID   FRE6_YEAST     STANDARD;      PRT;   712 AA.
AC   Q12473;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Ferric reductase transmembrane component 6 precursor (EC 1.6.99.13)
DE   (Ferric-chelate reductase 6).
GN   FRE6 OR YLL051C OR L0593.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c;
RA   Wedler H., Wambutt R.;
RL   Submitted (JAN-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: NADH + 2 Fe(3+) = NAD(+) + 2 Fe(2+).
CC   -!- COFACTOR: FAD (PROBABLE).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: BELONGS TO THE FRE / CYBB FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z47973; CAA88006.1; -.
DR   EMBL; Z73156; CAA97503.1; -.
DR   SGD; S0003974; FRE6.
DR   InterPro; IPR002916; Ferric_reduct.
DR   Pfam; PF01794; Ferric_reduct; 1.
KW   Oxidoreductase; Electron transport; Transmembrane; Iron transport;
KW   FAD; NAD; Glycoprotein; Signal; Multigene family.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18    712       FERRIC REDUCTASE TRANSMEMBRANE COMPONENT
FT                                6.
FT   NP_BIND     493    499       FAD (POTENTIAL).
FT   TRANSMEM    168    188       POTENTIAL.
FT   TRANSMEM    245    265       POTENTIAL.
FT   TRANSMEM    288    308       POTENTIAL.
FT   TRANSMEM    329    349       POTENTIAL.
FT   TRANSMEM    361    381       POTENTIAL.
FT   TRANSMEM    388    408       POTENTIAL.
FT   TRANSMEM    417    437       POTENTIAL.
FT   TRANSMEM    478    498       POTENTIAL.
FT   TRANSMEM    550    570       POTENTIAL.
FT   DOMAIN      553    556       POLY-LEU.
FT   CARBOHYD     89     89       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    112    112       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    124    124       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    450    450       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    655    655       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   712 AA;  81989 MW;  5224F12B51544BAA CRC64;

  Query Match             4.6%;  Score 109;  DB 1;  Length 712;
  Best Local Similarity   18.8%;  Pred. No. 2;
  Matches   72;  Conservative   51;  Mismatches  119;  Indels  142;  Gaps   17;

Qy      88 IIFVAIHREHYTSL-----------WDLRHLLVGKILIDVSNNMRINQYPESNAEYLASL 136
           || |  |  ||  |            :   | ::   |:|       | :   | |    |
Db     183 IIAVFFHMSHYNGLNRALFASRFVNYIRGHFVLPTFLVD----KHANHFKFLNVEVFTGL 238

Qy     137 FPDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIELARQLNFIPIDLGSLS 196
            |:||        ||
Db     239 MPNSL--------EAW------------------------------------------- 246

Qy     197 SAREIENLPLRLFTLWRGPVVVAISLATFFFL-YSFVRDVIHPYARNQQSDFYKIPIEIV 255
                         ::   :||   ||  |:: |  :    :  | | ::   :
Db     247 -------------------IIFGYTLANIIFLSISYIIDPYNLIFNSHLSQFTRL---LA 284

Qy     256 NKTLPIVAITLLSLVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQCRKQLGLLSFFFAMV 315
           :::  |:| |   |: :     :  :   | |:  |       :   | :| :    | :
Db     285 DRS-GILAFTQFPLIIIFTARNSFLEFLTGVKFNSF------ISFHKWIGRIMVLNATI 336

Qy     316 H-VAYSLCLPMRRSERYLFLNMAYQQVHANIENSWNEEEVWRIEMYISFGIMSLGLLSLL 374
           | ::|||         :  :| |::    | |          ::|  ||| |: :| :|
Db     337 HSLSYSL---------FAIINHAFK-----ISNK---------QLYWKFGIASITVLCVL 373

Qy     375 AVTSIPSVSNALNWREFSFIQSTLGYVALLI---STFHVLIY-GWKRAFE--------EE 422
           | |:  |      | : |   |   :|||       || |: |||            |:
```

```
Db    374 LVLSLGIVRK----RHYEFFLYTHIILALLFFYCCWQHVKIFNGWKEWIVVSLLIWGLEK 429

Qy    423 YYRFYTPPNFVLALVLPSIVILDL 446
          :| :     :|    |   :::|
Db    430 LFRIWN----ILQFRFPKATLINL 449


RESULT 8
FRE7_YEAST
ID   FRE7_YEAST     STANDARD;      PRT;   629 AA.
AC   Q12333;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Ferric reductase transmembrane component 7 (EC 1.6.99.13) (Ferric-
DE   chelate reductase 7).
GN   FRE7 OR YOL152W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / FY1679;
RX   MEDLINE=96132030; PubMed=8553699;
RA   Casamayor A., Aldea M., Casas C., Herrero E., Gamo F.J.,
RA   Lafuente M.J., Gancedo C., Arino J.;
RT   "DNA sequence analysis of a 13 kbp fragment of the left arm of yeast
RT   chromosome XV containing seven new open reading frames.";
RL   Yeast 11:1281-1288(1995).
CC   -!- CATALYTIC ACTIVITY: NADH + 2 Fe(3+) = NAD(+) + 2 Fe(2+).
CC   -!- COFACTOR: FAD (PROBABLE).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: BELONGS TO THE FRE / CYBB FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z48239; CAA88276.1; -.
DR   EMBL; Z74894; CAA99174.1; -.
DR   SGD; S0005512; FRE7.
DR   InterPro; IPR002916; Ferric_reduct.
DR   Pfam; PF01794; Ferric_reduct; 1.
KW   Oxidoreductase; Electron transport; Transmembrane; Iron transport;
KW   FAD; NAD; Glycoprotein; Multigene family.
FT   NP_BIND     369    375       FAD (POTENTIAL).
FT   TRANSMEM     46     66       POTENTIAL.
FT   TRANSMEM    108    128       POTENTIAL.
FT   TRANSMEM    168    188       POTENTIAL.
FT   TRANSMEM    195    215       POTENTIAL.
FT   TRANSMEM    238    258       POTENTIAL.
FT   TRANSMEM    266    286       POTENTIAL.
FT   TRANSMEM    293    313       POTENTIAL.
FT   TRANSMEM    422    442       POTENTIAL.
FT   CARBOHYD    330    330       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    541    541       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   629 AA;  71996 MW;  2384480E9289C16F CRC64;

  Query Match             4.6%;  Score 107.5;  DB 1;  Length 629;
  Best Local Similarity   21.7%;  Pred. No. 2.2;
  Matches   57;  Conservative   46;  Mismatches   77;  Indels   83;  Gaps   17;

Qy    214 GPVVVAISLATFFFLYSFVRDVIHPYARNQQSDFYKIPIEIVNKTLPIVAITLLSLVY-L 272
          |  :| ::   :  || ||    ||: |   : |   |: :      |:||:|:  |: |
Db    123 GTFLVVMATTLYTLLYCFVP---HPFYR-PCAGFGSPPLSV---RAGIMAISLVPFVFSL 175

Qy    273 AG-------LLAAAYQLYYGTKYRRFPPWLETWLQCRKQLGLLSFFFAMVHVAYSLCLPM 325
          :|       |:  :|:     |   :  |           :|  ||: |||     :|

Db    176 SGKINVIGWLVGLSYE-----KINIYHQW----------ASILCLFFSWVHV-----IPF 215

Qy    326 RRSERYLFLNMAYQQVHANIENSWNEEEVWRIEMYISFGIMSLGLLSLLAVTSIP----- 380
            |  |:       |:::|      |   ::||       |:  :  |:|| ::|:|
Db    216 LRQARH---EGGYERMH----QRWKASDMWR------SGVPPILFLNLLLWLSSLPIARRH 262

Qy    381 --SVSNALNWREFSFIQSTLGYVALLISTFHV-----------LIYGWKRAFEEEYYR- 425
            :   |:|     | :   |::|    :||             |  |   | : :||
Db    263 FYEIFLQLHW-----ILAVGFYISLF---YHVYPELNSHMYLVATIVVW---FAQLFYRL 311

Qy    426 ----FYTPPNFVLALVLPSIVIL 444
              :  |    :|  : :: |:
Db    312 AVKGYLRPGRSFMASTIANVSIV 334


RESULT 9
FSHR_HUMAN
ID   FSHR_HUMAN     STANDARD;      PRT;   695 AA.
AC   P23945;
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Follicle stimulating hormone receptor precursor (FSH-R) (Follitropin
DE   receptor).
GN   FSHR.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Ovary;
RX   MEDLINE=91222171; PubMed=1709010;
RA   Minegish T., Nakamura K., Takakura Y., Ibuki Y., Igarashi M.;
RT   "Cloning and sequencing of human FSH receptor cDNA.";
RL   Biochem. Biophys. Res. Commun. 175:1125-1130(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RX   MEDLINE=93246012; PubMed=1301382;
RA   Kelton C.A., Cheng S.V., Nugent N.P., Schweickhardt R.L.,
RA   Rosenthal J.L., Overton S.A., Wands G.D., Kuzeja J.B., Luchette C.A.,
RA   Chappel S.C.;
RT   "The cloning of the human follicle stimulating hormone receptor and
RT   its expression in COS-7, CHO, and Y-1 cells.";
RL   Mol. Cell. Endocrinol. 89:141-151(1992).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Tilly L.T., Aihara T., Nishimori K., Jai X.-C., Billig H.,
RA   Kowalski K.I., Perlas E.A., Hsueh A.J.;
RL   Submitted (XXX-1992) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 1-342 FROM N.A.
RC   TISSUE=Testis;
RX   MEDLINE=93075197; PubMed=1359889;
RA   Gromoll J., Gudermann T., Nieschlag E.;
RT   "Molecular cloning of a truncated isoform of the human follicle
RT   stimulating hormone receptor.";
RL   Biochem. Biophys. Res. Commun. 188:1077-1083(1992).
RN   [5]
RP   SEQUENCE OF 1-51 FROM N.A.
RX   MEDLINE=95011044; PubMed=7926278;
RA   Gromoll J., Dankbar B., Gudermann T.;
RT   "Characterization of the 5' flanking region of the human follicle-
RT   stimulating hormone receptor gene.";
RL   Mol. Cell. Endocrinol. 102:93-102(1994).
RN   [6]
RP   3D-STRUCTURE MODELING OF 49-228.
RX   MEDLINE=96363672; PubMed=8747461;
RA   Jiang X., Dreano M., Buckler D.R., Cheng S., Ythier A., Wu H.,
RA   Hendrickson W.A., el Tayar N.;
RT   "Structural predictions for the ligand-binding region of glycoprotein
RT   hormone receptors and the nature of hormone-receptor interactions.";
```

```
RL   Structure 3:1341-1353(1995).
CC   -!- FUNCTION: RECEPTOR FOR FOLLICLE STIMULATING HORMONE. THE ACTIVITY
CC       OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE
CC       ADENYLATE CYCLASE.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC       SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: SERTOLI CELLS AND OVARIAN GRANULOSA CELLS.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC       FSH/LSH/TSH SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 5 LEUCINE-RICH REPEATS (LRR).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M65085; AAA52477.1; -.
DR   EMBL; S59900; AAB26480.1; -.
DR   EMBL; M95489; AAA52478.1; -.
DR   EMBL; X68044; CAA48179.1; -.
DR   EMBL; S73199; AAB32071.1; -.
DR   PIR; JN0122; JN0122.
DR   PDB; 1XUN; 15-MAY-97.
DR   Genew; HGNC:3969; FSHR.
DR   MIM; 136435; -.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000372; LRR_Nterm.
DR   Pfam; PF00001; 7tm_1; 1.
DR   Pfam; PF00560; LRR; 3.
DR   Pfam; PF01462; LRRNT; 1.
DR   SMART; SM00013; LRRNT; 1.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
KW   Phosphorylation; Repeat; Leucine-rich repeat; Alternative splicing;
KW   3D-structure.
FT   SIGNAL       1     17       POTENTIAL.
FT   CHAIN       18    695       FOLLICLE STIMULATING HORMONE RECEPTOR.
FT   DOMAIN      18    366       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   367    387       1 (POTENTIAL).
FT   DOMAIN     388    398       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   399    421       2 (POTENTIAL).
FT   DOMAIN     422    443       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   444    465       3 (POTENTIAL).
FT   DOMAIN     466    485       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   486    508       4 (POTENTIAL).
FT   DOMAIN     509    528       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   529    550       5 (POTENTIAL).
FT   DOMAIN     551    573       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   574    597       6 (POTENTIAL).
FT   DOMAIN     598    608       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   609    630       7 (POTENTIAL).
FT   DOMAIN     631    695       CYTOPLASMIC (POTENTIAL).
FT   REPEAT      69     93       LRR 1.
FT   REPEAT     119    143       LRR 2.
FT   REPEAT     170    192       LRR 3.
FT   REPEAT     193    216       LRR 4.
FT   REPEAT     218    240       LRR 5.
FT   DISULFID   442    517       BY SIMILARITY.
FT   CARBOHYD   191    191       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   199    199       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   293    293       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   318    318       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC   224    285       MISSING (IN SHORT ISOFORM).
FT   VARSPLIC   342    695       MISSING (IN SHORT ISOFORM).
FT   CONFLICT    13     13       S -> R (IN REF. 4).
FT   CONFLICT   112    112       N -> T (IN REF. 1).
FT   CONFLICT   197    198       EL -> AV (IN REF. 1).
FT   CONFLICT   295    295       S -> P (IN REF. 4).
FT   CONFLICT   307    307       T -> A (IN REF. 1).
FT   CONFLICT   680    680       N -> S (IN REF. 1).
SQ   SEQUENCE   695 AA;  78294 MW;  723B8E71F76D2CD5 CRC64;

  Query Match             4.6%;  Score 107.5;  DB 1;  Length 695;
  Best Local Similarity   17.0%;  Pred. No. 2.5;
  Matches   84;  Conservative   77;  Mismatches  148;  Indels  185;  Gaps   17;

Qy      79 HEDALTKTNIIFVAIHREHYTSLWDLRHLLVG----------------KILIDVSNNMR 121
           ||  : | | : : |: | : :| :|::||:                  |:|:|: :|:
Db      98 HEIRIEKANNL-LYINPEAFQNLPNLQYLLISNTGIKHLPDVHKIHSLQKVLLDIQDNIN 156

Qy     122 INQYPESNAEYLASLFPDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIEL 181
           |:          :: :| :   |  |  |      |:  :::: |:
Db     157 IH----------TIERNSFVGLSFESVILWL----NKNGIQEIHNCA------------ 189

Qy     182 ARQLNFIPIDLGSLSSAREIENLPLRLFTLWRGPVVVAISLATFFFLYSFVRDVIHPYAR 241
               |   :|  :||    :| ||  :|    |||:: ||      | |:  : :
Db     190 ---FNGTQLDELNLSDNNNLEELPNDVFHGASGPVILDISRTRIHSLPSYGLENLKKLRA 246

Qy     242 NQQSDFYKIPIEIVNKTL-PIVAITLLSLVY---------------------------- 271
           :  |:|      ||  :||:   || |
Db     247 RSTYNLKKLP------TLEKLVALMEASLTYPSHCCAFANWRRQISELHPICNKSILRQE 300

Qy     272 -------------LAGLLAAAYQLYYGTKYRRF-------------PPWLETWLQCRKQL 305
                        ||    ::|   :   |  |             |  : :  |   :
Db     301 VDYMTQTRGQRSSLAEDNESSYSRGFDMTYTEFDYDLCNEVVDVTCSPKPDAFNPCEDIM 360

Qy     306 G-----LLSFFFAMVHVA-------------YSLCLPMRRSERYLFLNMAYQQVHANIEN 347
           |     :| :| ::: :               | | :|     |:|  |:|:
Db     361 GYNILRVLIWFISILAITGNIIVLVILTTSQYKLTVP-----RFLMCNLAFAD------- 408

Qy     348 SWNEEEVWRIEMYISFGIMSLGLLSLLAVTSIPSVSNALNWRE---------FSFIQSTL 398
                       :  ||  | : |:   |     : |::|:          |:   | |
Db     409 -------------LCIGIYLLLIASVDIHTKSQYHNYAIDWQTGAGCDAAGFFTVFASEL 455

Qy     399 G---YVALLISTFH------------------VLIYGWKRAFEEEYYRFYTPPNFV-LA 435
               |: :  :|                    |:: ||  ||    :  :   ::: ::
Db     456 SVYTLTAITLERWHTITHAMQLDCKVQLRHAASVMVMGWIFAFAAALFPIFGISSYMKVS 515

Qy     436 LVLPSIVILDLLQL 449
           : ||  :   | ||
Db     516 ICLPMDIDSPLSQL 529


RESULT 10
FSHR_PIG
ID   FSHR_PIG       STANDARD;      PRT;   695 AA.
AC   P49059; O77514;
DT   01-FEB-1996 (Rel. 33, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Follicle stimulating hormone receptor precursor (FSH-R) (Follitropin
DE   receptor).
GN   FSHR.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Ovary;
RX   MEDLINE=96011644; PubMed=7590277;
RA   Remy J.J., Lahbib-Mansais Y., Yerle M., Bozon V., Couture L.,
RA   Pajot E., Grebert D., Salesse R.;
RT   "The porcine follitropin receptor: cDNA cloning, functional
RT   expression and chromosomal localization of the gene.";
RL   Gene 163:257-261(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Ovary;
RA   Wang Y.F., Meyer K.B., Schmidt K., Wan S.J., Degen S.J.F.,
```

```
RA   la Barbera A.R.;
RT   "Porcine follicle-stimulating hormone receptor.";
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: RECEPTOR FOR FOLLICLE STIMULATING HORMONE. THE ACTIVITY
CC       OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE
CC       ADENYLATE CYCLASE.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC       FSH/LSH/TSH SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 6 LEUCINE-RICH REPEATS (LRR).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L31966; AAA86933.1; -.
DR   EMBL; AF025377; AAC24981.1; -.
DR   HSSP; P23945; 1XUN.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000372; LRR_Nterm.
DR   Pfam; PF00001; 7tm_1; 1.
DR   Pfam; PF00560; LRR; 2.
DR   Pfam; PF01462; LRRNT; 1.
DR   SMART; SM00013; LRRNT; 1.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
KW   Phosphorylation; Repeat; Leucine-rich repeat.
FT   SIGNAL       1    17       POTENTIAL.
FT   CHAIN       18   695       FOLLICLE STIMULATING HORMONE RECEPTOR.
FT   DOMAIN      18   366       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   367   387       1 (POTENTIAL).
FT   DOMAIN     388   398       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   399   421       2 (POTENTIAL).
FT   DOMAIN     422   443       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   444   465       3 (POTENTIAL).
FT   DOMAIN     466   485       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   486   508       4 (POTENTIAL).
FT   DOMAIN     509   528       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   529   550       5 (POTENTIAL).
FT   DOMAIN     551   573       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   574   597       6 (POTENTIAL).
FT   DOMAIN     598   608       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   609   630       7 (POTENTIAL).
FT   DOMAIN     631   695       CYTOPLASMIC (POTENTIAL).
FT   REPEAT      44    68       LRR 1.
FT   REPEAT      69    93       LRR 2.
FT   REPEAT     119   143       LRR 3.
FT   REPEAT     170   192       LRR 4.
FT   REPEAT     193   216       LRR 5.
FT   REPEAT     218   240       LRR 6.
FT   DISULFID   442   517       BY SIMILARITY.
FT   CARBOHYD   191   191       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   199   199       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   293   293       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT     2     2       S -> A (IN REF. 1).
FT   CONFLICT    13    13       T -> S (IN REF. 1).
FT   CONFLICT    60    60       V -> A (IN REF. 1).
FT   CONFLICT   166   166       V -> M (IN REF. 1).
FT   CONFLICT   215   215       Q -> H (IN REF. 1).
FT   CONFLICT   247   247       K -> R (IN REF. 1).
FT   CONFLICT   257   257       S -> T (IN REF. 1).
FT   CONFLICT   334   334       D -> N (IN REF. 1).
FT   CONFLICT   349   349       E -> K (IN REF. 1).
FT   CONFLICT   352   352       T -> A (IN REF. 1).
FT   CONFLICT   383   383       V -> E (IN REF. 1).
FT   CONFLICT   407   407       A -> T (IN REF. 1).
FT   CONFLICT   421   421       V -> I (IN REF. 1).
FT   CONFLICT   427   427       T -> S (IN REF. 1).
FT   CONFLICT   435   435       D -> N (IN REF. 1).
FT   CONFLICT   483   483       L -> V (IN REF. 1).
FT   CONFLICT   550   550       T -> I (IN REF. 1).
FT   CONFLICT   586   586       A -> V (IN REF. 1).
FT   CONFLICT   607   607       S -> L (IN REF. 1).
FT   CONFLICT   691   691       R -> H (IN REF. 1).
SQ   SEQUENCE   695 AA;  78172 MW;  E9EBEDB29C79C450 CRC64;
```

```
  Query Match             4.4%;  Score 104.5;  DB 1;  Length 695;
  Best Local Similarity   16.5%;  Pred. No. 4.1;
  Matches   91;  Conservative   91;  Mismatches  169;  Indels  199;  Gaps   21;

Qy      18 LPNG-INGIKDARKVTVGVIGSGDFAKSLTIRLIRCGYHVVIGSRNPKFASEFFPHVVDV 76
           :| |  :|  |  |:         : :::  : :|       : |  ||
Db      61 IPKGAFSGFGDLEKI--------EISQNDVLEVIEAN----VFSNLPKL---------- 97

Qy      77 THHEDALTKTNIIFVAIHREHYTSLWDLRHLLVG----------------KILIDVSNN 119
             ||  : | | : : |  : : :| :||:||:          .       |:|:|: :|
Db      98 --HEIRIEKANNL-LYIDPDAFQNLPNLRYLLISNTGVKHLPAVHKIQSLQKVLLDIQDN 154

Qy     120 MRINQYPESNAEYLASLFPDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVI 179
           : |:          ::  :| :  |  :  |      |:  |::: |:
Db     155 INIH----------TVERNSFVGLSFESMILWL----SKNGIREIHNCA---------- 189

Qy     180 ELARQLNFIPIDLGSLSSAREIENLPLRLFTLWRGPVVVAISLATFFFLYSFVRDVIHPY 239
             |   :|  :||    :| ||  :|    |||:: ||      | |:  : :
Db     190 -----FNGTQLDELNLSDNDNLEELPNDVFQGASGPVILDISRTRIHSLPSYGLENLKKL 244

Qy     240 ARNQQSDFYKIP-----IEIVNKTL-------------------PIVAITLL------- 267
             :  |:|     : ::  :|                    . ||   ::|
Db     245 RAKSTYNLKKLPSLEKFVTLMEASLTYPSHCCAFANWRRQISDLHPICNKSILRQEVDVM 304

Qy     268 -----SLVYLAGLLAAAYQLYYGTKYRRF------------PPWLETWLQCRKQLG--- 306
               | ||    ::   : | |  |                 |  :|:  |   :|
Db     305 TQARGQRVSLAEDGESSLAKEFDTMYSEFDYDLCNEVVDVICSPEPDTFNPCEDIMGHDI 364

Qy     307 --LLSFFFAMVHVA------------YSLCLPMRRSERYLFLNMAYQQVHANIENSWNE 351
             :| :| ::: :               | | :|     |:|  |:|:
Db     365 LRVLIWFISILAITGNIIVLVILITSQYKLTVP-----RFLMCNLAFAD---------- 408

Qy     352 EEVWRIEMYISFGIMSLGLLSLLAVTSIPSVSNALNWRE---------FSFIQSTLG--- 399
               :  ||  | : |:   |      : |::|:          |:   | |
Db     409 ---------LCIGIYLLLIASVDIHTKTQYHNYAIDWQTGAGCDAAGFFTVFASELSVYT 459

Qy     400 YVALLISTFH------------------VLIYGWKRAFEEEYYRFYTPPNFV-LALVLP 439
             |: :  :|                   ::: ||  ||    :  :   ::: ::: ||
Db     460 LTAITLERWHTITHAMQLQCKVQLRHAASIMLVGWIFAFTVALFPIFGISSYMKVSICLP 519

Qy     440 SIVILDLLQL 449
             :   | ||
Db     520 MDIDSPLSQL 529
```

```
RESULT 11
SOTB_ECO57
ID   SOTB_ECO57     STANDARD;      PRT;   396 AA.
AC   P58529;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Sugar efflux transporter.
GN   SOTB OR Z2173 OR ECS2135.
OS   Escherichia coli O157:H7.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=83334;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11206551;
RA   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
```

```
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / RIMD 0509952;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
CC   -!- FUNCTION: Involved in the efflux of sugars. The physiological role
CC       may be the reduction of the intracellular concentration of toxic
CC       sugars or sugar metabolites. Transports L-arabinose and to a
CC       lesser extent IPTG. Seems to contribute to the control of the
CC       arabinose regulon (By similarity).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC       (Probable).
CC   -!- SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (ALSO KNOWN AS
CC       THE DRUG RESISTANCE TRANSLOCASE FAMILY). SOTB (TC 2.A.1.2)
CC       SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE005352; AAG56236.1; -.
DR   EMBL; AP002557; BAB35558.1; -.
KW   Transport; Sugar transport; Transmembrane; Inner membrane;
KW   Complete proteome.
FT   TRANSMEM     15     35       POTENTIAL.
FT   TRANSMEM     50     70       POTENTIAL.
FT   TRANSMEM     81    101       POTENTIAL.
FT   TRANSMEM    103    123       POTENTIAL.
FT   TRANSMEM    136    156       POTENTIAL.
FT   TRANSMEM    170    190       POTENTIAL.
FT   TRANSMEM    209    229       POTENTIAL.
FT   TRANSMEM    246    266       POTENTIAL.
FT   TRANSMEM    275    295       POTENTIAL.
FT   TRANSMEM    299    319       POTENTIAL.
FT   TRANSMEM    333    353       POTENTIAL.
FT   TRANSMEM    364    384       POTENTIAL.
SQ   SEQUENCE   396 AA;  42568 MW;  CB6A209B0BE6D4F0 CRC64;

  Query Match             4.4%;  Score 104;  DB 1;  Length 396;
  Best Local Similarity   18.8%;  Pred. No. 2.3;
  Matches   85;  Conservative   72;  Mismatches  148;  Indels  146;  Gaps   20;

Qy      31 VTVGVIGSGDFAKSLTIRLIRCGYHVVIGSRNPKFASEFFPHVVDVTHHEDALTKTNIIF 90
           | ||::   | |:|  ::  : |  : | :      |  |  :          |    ::|
Db      31 VPVGLL--SDIAQSFHMQTAQVGIMLTIYAWVVALMSLPFMLMTSQVERRKLLICLFVVF 88

Qy      91 VAIHREHYTSLWDLRHLLVGKILIDVSNNMRINQYPESNAEYLASLFPDSLIVKGFNVVS 150
           :| |   : | |    |:: :|          :  ::|        ::: :
Db      89 IASHVLSFLS-WSFTVLVISRI----------------GVAFAHAIF--------WSITA 123

Qy     151 AWALQLGP--KDASRQVYICSNNIQARQQVIELAR--------QLNFIPIDLGSL----- 195
           : |::: |  | |     | :    |    : | |        :: |  | :|:|
Db     124 SLAIRMAPAGKRAQALSLIATGTALAMVLGLPLGRIVGQYFGWRMTFFAIGIGALITLLC 183

Qy     196 ----------SSAREIENLPLRLFTLWRGPVVVAISLAT------FFFLYSFVRDVIHPY 239
                    :  :::|||    |:| | :::| | |       :  ||:    | |:
Db     184 LIKLLPLLPSEHSGSLKSLPL----LFRRPALMSIYLLTVVVVTAHYTAYSY----IEPF 235

Qy     240 ARN---QQSDFYKIPIEIVNKTLPIVAITLLSLVYLAGLLAAAYQLYYGTKYRRFPPWLE 296
           :|    ::|               |  || |:  ||:: :      | :|
Db     236 VQNIAGFSANF--------------ATALLLLGGAGIIGSVIFGKLGNQYAS------ 274

Qy     297 TWLQCRKQLGLLSFFFAMVHVAYSLCLPMRRSERYLFLNMAYQQVHANIENSWNEEEVWR 356
                 |:|   |:: |  :| ||   ||           :|  : : :     |
Db     275 ---------ALVSTAIALLLVCLALLLPAANSE----------IHLGVLSIF-----WG 309

Qy     357 IEMYISFGIMSLGLLSL------LAVTSIPSVSN-------------ALNWREFSFIQST 397
           | | |    | : :|:|      :|:     : |             :|:|       |
Db     310 IAMMIIGLGMQVKVLALAPDATDVAMALFSGIFNIGIGAGALVGNQVSLHWS-----MSM 364

Qy     398 LGYVALLISTFHVLIYG------WKRAFEEE 422
           :|||   :  |  ||:       |    ||:
Db     365 IGYVG-TVPAFAALIWSIIIFRRWPVTLEEQ 394


RESULT 12
SOTB_ECOLI
ID   SOTB_ECOLI     STANDARD;      PRT;   396 AA.
AC   P31122; P77353; P76883;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Sugar efflux transporter.
GN   SOTB OR B1528.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=97251357; PubMed=9097039;
RA   Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA   Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA   Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA   Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N.,
RA   Sampei G., Seki Y., Sivasundaram S., Tagami H., Takeda J.,
RA   Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
RT   "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 28.0-40.1 min region on the linkage map.";
RL   DNA Res. 3:363-377(1996).
RN   [3]
RP   SEQUENCE OF 217-396 FROM N.A.
RX   MEDLINE=93186717; PubMed=8383113;
RA   Cohen S.P., Haechler H., Levy S.B.;
RT   "Genetic and functional analysis of the multiple antibiotic
RT   resistance (mar) locus in Escherichia coli.";
RL   J. Bacteriol. 175:1484-1492(1993).
RN   [4]
RP   CHARACTERIZATION.
RC   STRAIN=SB0;
RX   MEDLINE=99194728; PubMed=10094697;
RA   Bost S., Silva F., Belin D.;
RT   "Transcriptional activation of ydeA, which encodes a member of the
RT   major facilitator superfamily, interferes with arabinose accumulation
RT   and induction of the Escherichia coli arabinose PBAD promoter.";
RL   J. Bacteriol. 181:2185-2191(1999).
RN   [5]
RP   CHARACTERIZATION.
RC   STRAIN=JS219;
```

```
RX   MEDLINE=99369894; PubMed=10438792;
RA   Carole S., Pichoff S., Bouche J.-P.;
RT   "Escherichia coli gene ydeA encodes a major facilitator pump which
RT   exports L-arabinose and isopropyl-beta-D-thiogalactopyranoside.";
RL   J. Bacteriol. 181:5123-5125(1999).
CC   -!- FUNCTION: Involved in the efflux of sugars. The physiological role
CC       may be the reduction of the intracellular concentration of toxic
CC       sugars or sugar metabolites. Transports L-arabinose and to a
CC       lesser extent IPTG. Seems to contribute to the control of the
CC       arabinose regulon.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC       (Probable).
CC   -!- SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (ALSO KNOWN AS
CC       THE DRUG RESISTANCE TRANSLOCASE FAMILY). SOTB (TC 2.A.1.2)
CC       SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000250; AAC74601.1; -.
DR   EMBL; D90795; BAA15210.1; -.
DR   EMBL; D90796; BAA15218.1; -.
DR   EMBL; D90797; BAA15230.1; -.
DR   EMBL; M96235; -; NOT_ANNOTATED_CDS.
DR   EcoGene; EG11636; sotB.
DR   InterPro; IPR003662; sub_transporter.
DR   Pfam; PF00083; sugar_tr; 1.
KW   Transport; Sugar transport; Transmembrane; Inner membrane;
KW   Complete proteome.
FT   TRANSMEM     15     35       POTENTIAL.
FT   TRANSMEM     50     70       POTENTIAL.
FT   TRANSMEM     81    101       POTENTIAL.
FT   TRANSMEM    103    123       POTENTIAL.
FT   TRANSMEM    136    156       POTENTIAL.
FT   TRANSMEM    170    190       POTENTIAL.
FT   TRANSMEM    209    229       POTENTIAL.
FT   TRANSMEM    246    266       POTENTIAL.
FT   TRANSMEM    275    295       POTENTIAL.
FT   TRANSMEM    299    319       POTENTIAL.
FT   TRANSMEM    333    353       POTENTIAL.
FT   TRANSMEM    364    384       POTENTIAL.
SQ   SEQUENCE   396 AA;  42538 MW;  CB6A34CA4EE6D4F0 CRC64;

  Query Match             4.4%;  Score 104;  DB 1;  Length 396;
  Best Local Similarity   18.8%;  Pred. No. 2.3;
  Matches   85;  Conservative   72;  Mismatches  148;  Indels  146;  Gaps   20;

Qy      31 VTVGVIGSGDFAKSLTIRLIRCGYHVVIGSRNPKFASEFFPHVVDVTHHEDALTKTNIIF 90
           | ||::   | |:|  ::  : |  : | :      |  |  :         |    ::|
Db      31 VPVGLL--SDIAQSFHMQTAQVGIMLTIYAWVVALMSLPFMLMTSQVERRKLLICLFVVF 88

Qy      91 VAIHREHYTSLWDLRHLLVGKILIDVSNNMRINQYPESNAEYLASLFPDSLIVKGFNVVS 150
           :| |   : | |    |:: :|                  :  ::|       ::: :
Db      89 IASHVLSFLS-WSFTVLVISRI---------------GVAFAHAIF--------WSITA 123

Qy     151 AWALQLGP--KDASRQVYICSNNIQARQQVIELAR--------QLNFIPIDLGSL----- 195
           : |::: |  | |    | :    |   : | |         :: |  | :|:|
Db     124 SLAIRMAPAGKRAQALSLIATGTALAMVLGLPLGRIVGQYFGWRMTFFAIGIGALITLLC 183

Qy     196 ----------SSAREIENLPLRLFTLWRGPVVVAISLAT------FFFLYSFVRDVIHPY 239
                    :  :::|||    |:| | :::| | |       :  ||:    | |:
Db     184 LIKLLPLLPSEHSGSLKSLPL----LFRRPALMSIYLLTVVVVTAHYTAYSY----IEPF 235

Qy     240 ARN---QQSDFYKIPIEIVNKTLPIVAITLLSLVYLAGLLAAAYQLYYGTKYRRFPPWLE 296
            :|    ::|               |  || |:  ||:: :      | :|
Db     236 VQNIAGFSANF--------------ATALLLLLGGAGIIGSVIFGKLGNQYAS------ 274

Qy     297 TWLQCRKQLGLLSFFFAMVHVAYSLCLPMRRSERYLFLNMAYQQVHANIENSWNEEEVWR 356
                  |:|   |:: |  :| ||   ||          :|  : : :     |
Db     275 ---------ALVSTAIALLLVCLALLLPAANSE----------IHLGVLSIF-----WG 309

Qy     357 IEMYISFGIMSLGLLSL------LAVTSIPSVSN------------ALNWREFSFIQST 397
           | | |    | : :|:|      :|:     : |            :|:|       |
Db     310 IAMMIIGLGMQVKVLALAPDATDVAMALFSGIFNIGIGAGALVGNQVSLHWS-----MSM 364

Qy     398 LGYVALLISTFHVLIYG------WKRAFEEE 422
           :|||   :  |  ||:       |    ||:
Db     365 IGYVG-AVPAFAALIWSIIIFRRWPVTLEEQ 394


RESULT 13
NPT1_MOUSE
ID   NPT1_MOUSE     STANDARD;      PRT;   465 AA.
AC   Q61983;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Renal sodium-dependent phosphate transport protein 1 (Sodium/phosphate
DE   cotransporter 1) (Na(+)/PI cotransporter 1) (Renal sodium-phosphate
DE   transport protein 1) (Renal Na(+)-dependent phosphate cotransporter
DE   1).
GN   SLC17A1 OR NPT1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Kidney;
RX   MEDLINE=95335846; PubMed=7611445;
RA   Chong S.S., Kozak C.A., Liu L., Kristjansson K., Dunn S.T.,
RA   Bourdeau J.E., Hughes M.R.;
RT   "Cloning, genetic mapping, and expression analysis of a mouse renal
RT   sodium-dependent phosphate cotransporter.";
RL   Am. J. Physiol. 268:F1038-F1045(1995).
CC   -!- FUNCTION: IMPORTANT FOR THE RESORPTION OF PHOSPHATE BY THE KIDNEY.
CC       MAY BE INVOLVED IN ACTIVELY TRANSPORTING PHOSPHATE INTO CELLS VIA
CC       NA+ COTRANSPORT IN THE RENAL BRUSH BORDER MEMBRANE.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: KIDNEY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X77241; CAA54459.1; -.
DR   MGD; MGI:103209; Slc17a1.
DR   InterPro; IPR004745; Pi_cotransport.
DR   TIGRFAMs; TIGR00894; 2A0114euk; 1.
KW   Transport; Symport; Sodium transport; Transmembrane; Glycoprotein.
FT   TRANSMEM     79     99       POTENTIAL.
FT   TRANSMEM    117    137       POTENTIAL.
FT   TRANSMEM    176    196       POTENTIAL.
FT   TRANSMEM    199    219       POTENTIAL.
FT   TRANSMEM    260    280       POTENTIAL.
FT   TRANSMEM    304    324       POTENTIAL.
FT   TRANSMEM    337    356       POTENTIAL.
FT   TRANSMEM    363    383       POTENTIAL.
FT   TRANSMEM    399    419       POTENTIAL.
FT   TRANSMEM    429    449       POTENTIAL.
FT   CARBOHYD     39     39       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     47     47       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     56     56       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   465 AA;  51589 MW;  C67EE25A2C291EEF CRC64;

  Query Match             4.4%;  Score 103;  DB 1;  Length 465;
  Best Local Similarity   18.7%;  Pred. No. 3.3;
```

```
  Matches   76;  Conservative   64;  Mismatches  142;  Indels  124;  Gaps   16;

Qy     110 GKILIDVSNNMRINQYPESNAEYLASLFPDSLIVKGFNVVSAWALQLGPKDA---SRQVY 166
           | ||  |   | : | |     ||: ::|   |:     :|:     | |  |   :  |
Db      79 GLILSSVFFGMVVVQAP---VGYLSGIYPMKRIIGSSLFLSSLMSLLIPPAAQVGAALVI 135

Qy     167 ICSNNIQARQQVIELARQLNFI----PIDLGSLSSAREIENLPLRLFTLWRGPVVV---- 218
           :|      |  :   :   ::    |:: | |:|      : |  | :  || :|
Db     136 VCRVLQGIAQGTVSTGQHEIWVKWAPPLERGRLTS------MTLSGFVM--GPFIVLLVS 187

Qy     219 --------------------AISLATFFFLYSFVRDVIHPYARNQQSDF---------- 247
                               :||: ||  :   :|  |||  : : |:
Db     188 GFICDLLGWPMVFYIFGIVGCVLSLSWFFLFFDDPKD--HPYMSSSEKDYIISSLMQQAS 245

Qy     248 ---YKIPIEIVNKTLPIVAITLLS---------------------------------- 268
                :||: : |:||: || | |
Db     246 SGRQSLPIKAMLKSLPLWAIILNSFAFIWSNSLLVTYTPTFISTVLHVNVRENGLLSSLP 305

Qy     269 --LVYLAGLLAAAYQLYYGTKYRRFPPWLETWLQCRKQLGLLSFFFAMVHVAYSLCLPMR 326
             | |: |:||     :: |:       : : :  ||    |  |  :: :   | |
Db     306 YLLAYICGILAGQMSDFFLTRK------IFSIVTVRKLFTTLGSFCPVIFIMCLLYLSYN 359

Qy     327 RSERYLFLNMAYQQVHANIENSWNEEEVWRIEM---YISF--GIMSL-----GLLSLLAV 376
               :|| :|    :: :  |:  : :  :::   |  |   : :|     ||:|
Db     360 FYSTVIFLTLA----NSTLSFSYCGQLINALDIAPRYYGFLKAVTALIGMFGGLISSTLA 415

Qy     377 TSIPSVSNALNWREFSFIQSTLGYVALLISTFHVL-----IYGWKR 417
           | :      | : ||: : :    |:   |: |     |  | :
Db     416 GLILNQDPEYAWHKISFLMAGINVTCLV---FYFLFAKGEIQDWAK 458


RESULT 14
Y304_BRUME
ID   Y304_BRUME     STANDARD;      PRT;   220 AA.
AC   Q8YD73;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical protein BMEII0304.
GN   BMEII0304.
OS   Brucella melitensis.
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Brucellaceae; Brucella.
OX   NCBI_TaxID=29459;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=16M / ATCC 23456 / Biotype 1;
RX   MEDLINE=20020109; PubMed=11756688;
RA   DelVecchio V.G., Kapatral V., Redkar R.J., Patra G., Mujer C., Los T.,
RA   Ivanova N., Anderson I., Bhattacharyya A., Lykidis A., Reznik G.,
RA   Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltsman E.,
RA   Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
RA   Haselkorn R., Kyrpides N., Overbeek R.;
RT   "The genome sequence of the facultative intracellular pathogen
RT   Brucella melitensis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: BELONGS TO THE UPF0191 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE009669; AAL53546.1; -.
KW   Hypothetical protein; Transmembrane; Complete proteome.
FT   TRANSMEM     20     39       POTENTIAL.
FT   TRANSMEM     54     72       POTENTIAL.
FT   TRANSMEM     85    104       POTENTIAL.
FT   TRANSMEM    124    146       POTENTIAL.
FT   TRANSMEM    153    175       POTENTIAL.
FT   TRANSMEM    179    198       POTENTIAL.
SQ   SEQUENCE   220 AA;  24815 MW;  182C0244743B17FA CRC64;

  Query Match             4.3%;  Score 102;  DB 1;  Length 220;
  Best Local Similarity   23.8%;  Pred. No. 1.6;
  Matches   43;  Conservative   25;  Mismatches   51;  Indels   62;  Gaps    7;

Qy     267 LSLVYLAGLLAAAYQLYYGT-------KYRRFPPWLETW---------------------- 298
           | |:| || : | :  | |           : |   |  |
Db      20 LWLLYTAGFVPAVWTFYLGATGQLGADPVKTFEHLLGLWALRFLILTLLVTPMRDLTGIT 79

Qy     299 -LQCRKQLGLLSFFFAMVHVAYSLCLPMRRSERYLFLNMAYQQVHANIENSWNEEEVWRI 357
            |: |: ||||:|::|::|                      |  :   :  |    :: |
Db      80 LLRYRRALGLLAFYYALMHFT-----------------TYMVLDQGLNLSAIITDIVR- 120

Qy     358 EMYISFGIMSLGLLSLLAVTSIPSVSNALNW------REFSFIQSTLGYVALLISTFHVL 411
             :|: |::|| ||  ||:||        ||      | :| :   | |:|:  |  | |
Db     121 RPFITIGMISLALLVPLALTS-------NNWSIRKLGRRWSSLHK-LVYIAIAGSAVHFL 172

Qy     412 I 412
           :
Db     173 M 173


RESULT 15
RF3_SACBA
ID   RF3_SACBA      STANDARD;      PRT;   476 AA.
AC   P05512;
DT   01-NOV-1988 (Rel. 09, Created)
DT   01-NOV-1988 (Rel. 09, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Maturase-like RF3 protein.
GN   RF3.
OS   Saccharomyces bayanus (Yeast) (Saccharomyces uvarum).
OG   Mitochondrion.
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4931;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87280035; PubMed=2440860;
RA   Seraphin B., Simon M., Faye G.;
RT   "The mitochondrial reading frame RF3 is a functional gene in
RT   Saccharomyces uvarum.";
RL   J. Biol. Chem. 262:10146-10153(1987).
CC   -!- SIMILARITY: THE RF3 PROTEINS ARE RELATED TO THE MATURASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J03300; AAA32166.2; -.
DR   PIR; A28439; A28439.
DR   InterPro; IPR001982; Endonuc_LAG/HNH.
DR   Pfam; PF00961; LAGLIDADG_1; 1.
KW   Mitochondrion.
SQ   SEQUENCE   476 AA;  57863 MW;  ECB416C51DFFDA6D CRC64;

  Query Match             4.3%;  Score 100;  DB 1;  Length 476;
  Best Local Similarity   20.7%;  Pred. No. 5.5;
  Matches   52;  Conservative   48;  Mismatches   85;  Indels   66;  Gaps   12;

Qy      62 NPKFASEFFPHVVDVTHHEDALTKTNI----IFVAIHREHYT-------------SLWD 103
           || | : |  ::           |||:    ||  |: : |:             :
Db     254 NPYFVNAFSINI----------KTNLAKEKIFTNIYNKLYSDYKINQINNHIPYYNYLK 302

Qy     104 LRHLLVGKILIDVSNNMRINQYPESNAEYLASLF--PDSLIVKGFNVVSAWALQLGPKDA 161
           : : |  | ::|: ||  :  :  ::  :|:|::   |:|: |
```

```
Db    303 INNKLPIKNIMDIKNNYWLAGFTAADGSFLSSMYNPKDTLLFKDM---------------- 347

Qy    162 SRQVYICSNNIQARQQVIELARQLNFIPIDLGSLSSAREIENLPLRLFTLWRGPVVVAIS 221
          |  |: |  :: |:::| | ::      || |:|: ::: |  |: | |:       :
Db    348 -RPSYVIS-QVETRKELIYLIQE----SFDL-SISNVKKVGNRKLKDFKLFTRTTDELMK 400

Qy    222 LATFFFLYSFVRDVIHPYARNQQSDFYKIP----IEIVNKTLPIVAITLLSLVYLAGLLA 277
            |:| |  |   |   |:| :: |      |:  |    :  : |    |:  :
Db    401 -----FIYYF--DKFLPLHDNKQFNYIKFRFNTFIKSYNWNNRVFGLVLSE--YINNIKI 451

Qy    278 AAYQLYYGTKY 288
            |  ||  ||
Db    452 DNYDYYYYNKY 462

Search completed: January 17, 2003, 19:01:20
Job time : 20 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: January 17, 2003, 18:58:14 ; Search time 21 Seconds
(without alignments)
2078.335 Million cell updates/sec

Title: US-09-455-486-6
Perfect score: 2351
Sequence: 1 MESISMMGSPKSLSETCLPN.....ALVLPSIVILDLLQLCRYPD 454

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	160	6.8	198	2	D95285	conserved hypothet
2	159.5	6.8	211	2	AC2560	hypothetical prote
3	156.5	6.7	239	2	T50571	probable oxidoredu
4	147	6.3	213	2	H69400	conserved hypothet
5	142	6.0	212	2	D69361	conserved hypothet
6	142	6.0	224	2	T10120	F420-dependent NAD
7	133	5.7	223	2	D64487	hypothetical prote
8	121	5.1	191	2	F86826	hypothetical prote
9	120.5	5.1	232	2	A69131	conserved hypothet
10	119.5	5.1	216	2	T00121	hypothetical prote
11	116.5	5.0	242	2	G82642	conserved hypothet
12	116	4.9	198	2	AB3182	conserved hypothet
13	111.5	4.7	695	1	JN0898	follitropin recept
14	111.5	4.7	1228	2	S59681	probable membrane
15	110.5	4.7	222	2	B84410	hypothetical prote
16	110.5	4.7	694	2	JC4301	follitropin recept
17	109	4.6	320	2	T28379	ORF MSV218 hypothe
18	109	4.6	712	2	S50969	probable membrane
19	109	4.6	1184	2	H71436	hypothetical prote
20	109	4.6	1301	2	D85188	disease resistance
21	107.5	4.6	629	2	S60385	probable membrane
22	107.5	4.6	695	1	QRHUFT	follitropin recept
23	105	4.5	442	2	B64582	sodium- and chlori
24	105	4.5	604	2	T31042	hypothetical prote
25	104	4.4	396	1	C64907	chloramphenicol re
26	104	4.4	396	2	G90895	hypothetical prote
27	104	4.4	396	2	H85721	probable resistanc
28	103	4.4	465	2	S69915	sodium-phosphate t
29	102.5	4.4	314	2	H98310	probable peptide A
30	102.5	4.4	314	2	AB2972	hypothetical prote
31	102.5	4.4	1242	2	T39453	probable mrna stab
32	102	4.3	220	2	AG3547	bicyclomycin resis
33	102	4.3	574	2	T41068	hypothetical prote
34	101	4.3	320	2	E71139	hypothetical prote
35	101	4.3	348	2	T12284	NADH2 dehydrogenas
36	101	4.3	420	2	F69144	O-antigen transpor
37	101	4.3	501	2	T02134	hypothetical prote
38	101	4.3	503	2	C86250	hypothetical prote
39	101	4.3	735	2	A83006	hypothetical prote
40	100.5	4.3	346	2	T11181	NADH2 dehydrogenas
41	100.5	4.3	1780	2	A85045	probable glucan sy
42	100	4.3	395	2	C71219	hypothetical prote
43	100	4.3	476	2	A28439	endonuclease SceI
44	100	4.3	714	2	AF2479	ABC transporter AT
45	99.5	4.2	452	2	C71391	NADH2 dehydrogenas

ALIGNMENTS

RESULT 1
D95285
conserved hypothetical protein SMa0349 [imported] - Sinorhizobium meliloti (strain 10
C;Species: Sinorhizobium meliloti
C;Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C;Accession: D95285
R;Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; B
.; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K
Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001
A;Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meli
A;Reference number: A95262; MUID:21396509; PMID:11481432
A;Accession: D95285
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-198 <KUR>
A;Cross-references: GB:AE006469; PIDN:AAK64846.1; PID:g14523260; GSPDB:GN00165
A;Experimental source: strain 1021, megaplasmid pSymA
R;Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubl
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelau
hebault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh,
A;Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A;Reference number: A96039; MUID:21368234; PMID:11474104
A;Contents: annotation
C;Genetics:
A;Gene: SMa0349
A;Genome: plasmid

Query Match 6.8%; Score 160; DB 2; Length 198;
Best Local Similarity 26.7%; Pred. No. 2.1e-05;
Matches 55; Conservative 40; Mismatches 83; Indels 28; Gaps 8;

```
Qy      32 TVGVIGSGDFAKSLTIRLIRCGYHVVI-GSRNPKFAS---EFFPHVVDVTHHEDALTKTN 87
           |  :||:|    :|  |        :|  || |   |   : |   |     :||| : :
Db       3 TYAIIGAGAIGSALAERFTAAQIPAIIANSRGPASLSSVTDRFGASVKAVELKDAL-QAD 61

Qy      88 IIFVAIHREHYTSLWDLRHLLV---GKILIDVSNNMRINQYP------ESNAEYLASLFP 138
           :: :|:    | |: |:   :    |:|::| || :    :         : | :: | |
Db      62 VVILAV---PYDSIADIVTQVSDWGGQIVVDASNAIDFPAFKPRDLGGRLSTEIVSELVP 118

Qy     139 DSLIVKGFNVVSAWALQLGPK--DASRQVYICSNNIQARQQVIELARQLNFIPIDLGSLS 196
           : :|| || : |  |   |     || ::  |:  | :|| ||   | | |:|||:|:
Db     119 GAKVVKAFNTLPAAVLAADPDKGTGSRVLFLSGNHSDANRQVAELISSLGFAPVDLGTLA 178

Qy     197 SAREIENLPLRLFTLWRGPVVVAISL 222
           ::  |:          |  :||::|
Db     179 ASGPIQQF---------GRPLVALNL 195
```

RESULT 2

AC2560
hypothetical protein alr8074 [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120ga
C;Species: Nostoc sp.
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C;Accession: AC2560
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AC2560
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-211 <KUR>
A;Cross-references: GB:AP003603; PIDN:BAB77404.1; PID:g17134847; GSPDB:GN00182
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr8074
A;Genome: plasmid

Query Match 6.8%; Score 159.5; DB 2; Length 211;
Best Local Similarity 24.9%; Pred. No. 2.4e-05;
Matches 51; Conservative 43; Mismatches 70; Indels 41; Gaps 8;

```
Qy      33 VGVIGSGDFAKSLTIRLIRCGYHVVIGSRNP-----KFASEFFPHVVDVTHHEDALTKTN 87
           :  || |    :|  :|:   : | | :||       | |   :|  : |:  ::|: :
Db       3 IAFIGIGQVGSALASQLLSLDHTVTIAARNSNSDSVKTALAKYPE-LQVSSPQEAIAQAE 61

Qy      88 IIFVA----IHREHYTSLWDLRHLLVGKILIDVSNNMRIN-----QYPESNAEYLASLFP 138
           :||:|    ::      : ||     ||||:| :| : :|      :  :| :| : |  |
Db      62 VIFLATPFTANQAALAEVGDLS----GKILVDCTNPVGVNLTHGLKSEQSGSELVQSFVP 117

Qy     139 DSLIVKGFNVVSAWALQLGPKDASRQVY----------ICSNNIQARQQVIELARQLNF 187
           : :|| | :        | ::     |           |  |:: |:| |  |  :|| :
Db     118 HAKVVKAFTI-------YGFENFEHNTYPGYGNLKPAMLIAGNDVPAKQVVSTLCQQLGW 170

Qy     188 IPIDLGSLSSAREIENLPLRLFTLW 212
           :|:|:|| :  :|:: |    ||
Db     171 EAVDVGNLSMSLHLEHMTL----LW 191
```

RESULT 3
T50571
probable oxidoreductase [imported] - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 28-Jul-2000
C;Accession: T50571
R;Redenbach, M.; Kieser, H.M.; Denapaite, D.; Eichner, A.; Cullum, J.; Kinashi, H.; Hopw
Mol. Microbiol. 21, 77-96, 1996
A;Title: A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb S
A;Reference number: Z20556; MUID:97000351; PMID:8843436
A;Accession: T50571
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-239 <RED>
A;Cross-references: EMBL:AL133220; PIDN:CAB61708.1
A;Experimental source: strain A3(2)
C;Genetics:
A;Note: SCC75A.08c
C;Superfamily: conserved hypothetical protein MJ1501

Query Match 6.7%; Score 156.5; DB 2; Length 239;
Best Local Similarity 27.9%; Pred. No. 5e-05;
Matches 61; Conservative 41; Mismatches 72; Indels 45; Gaps 10;

```
Qy      25 IKDARKVTVGVI-GSGDFAKSLTIRLIRCGYHVVIGSR---NPKFASEFFPHVVDVTHHE 80
           :  |   : |||: |:|   | |  || : |  |::|||       |:|   | |:   :
Db      23 LPDVSGLVVGVLGGTGPQGKGLAYRLAKAGQKVIVGSRAAERAAAAAEEIGHGVEGADNA 82

Qy      81 DALTKTNIIFVAIHRE-HYTSLWDLRHLLVGKILIDVSNNMRINQ------YPE--SNAE 131
           :   :::::  ||:  : |  :|  ||  | ||:::|  | :  ::       ||  | ||
Db      83 ETARRSDVVIVAVPWDGHGKTLESLRAELSGKLVVDCVNPLGFDKKGAYALKPEEGSAAE 142
```

```
Qy     132 YLASLFPDSLIVKGFNVVSAWALQ-------------LGPKDASRQVYICSNNIQARQQV 178
           |:| |||  :   |: :||  ||              || : |  ::      :||
Db     143 QAAALLPDSRVAAAFHHLSAVLLQDPEIDEIDTDVMVLGEERADVEI------VQA---- 192

Qy     179 IELARQLNFIP----IDLGSLSSAREIENLPLRLFTLWR 213
             ||  :   ||    :  | | :| ::|:|   | :: |
Db     193 --LAGR---IPGMRGVFAGRLRNAHQVESLVANLISVNR 226
```

RESULT 4
H69400
conserved hypothetical protein AF1209 - Archaeoglobus fulgidus
C;Species: Archaeoglobus fulgidus
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 29-Sep-1999
C;Accession: H69400
R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A;Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes,
Smith, H.O.; Woese, C.R.; Venter, J.C.
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch
A;Reference number: A69250; MUID:98049343; PMID:9389475
A;Accession: H69400
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-213 <KLE>
A;Cross-references: GB:AE001021; GB:AE000782; NID:g2689344; PIDN:AAB90038.1; PID:g264
C;Superfamily: conserved hypothetical protein MJ1501

Query Match 6.3%; Score 147; DB 2; Length 213;
Best Local Similarity 27.5%; Pred. No. 0.00024;
Matches 56; Conservative 44; Mismatches 68; Indels 36; Gaps 11;

```
Qy      37 GSGDFAKSLTIRLIRCGYHVVIGSRN----PKFASEFFPHVVDVT----HHEDALTKTNI 88
           |:|:  : | :|  : || :::|||  \    | ||::   | | :     :|||    :
Db       8 GTGNLGEGLALRWGKLGYEIIVGSRKLEKAEKLASDYLKKVGDASIIGMRNEDAAETCD- 66

Qy      89 IFVAIHREHYTSLWD--------LRHLLVGKILID-------VSNNMRINQYPE--SNAE 131
             ||:    :|  |:          |:  | ||::|        | :|  :    ||  | ||
Db      67 --VAV----FTIPWEFAFDTAEMLKRQLAGKVVISPLVPMKKVGDNF-VYVRPEEGSAAE 119

Qy     132 YLASLFPDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIELARQLNFI-PI 190
           |||:  :| :|  :: : |        ::     | || ::  |:: |::|  ::: :  :
Db     120 KLASVLEESSVVAAYHSIPARRFANLGEEFEWDVPICGDS-GAKEVVVDLTEKISGLRAL 178

Qy     191 DLGSLSSAREIENL-PLRLFTLWR 213
           | | ||:|  :|:|  || |  : |
Db     179 DAGGLSNAHLVESLTPLILNVMKR 202
```

RESULT 5
D69361
conserved hypothetical protein AF0892 - Archaeoglobus fulgidus
C;Species: Archaeoglobus fulgidus
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 29-Sep-1999
C;Accession: D69361
R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A;Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes,
Smith, H.O.; Woese, C.R.; Venter, J.C.
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch
A;Reference number: A69250; MUID:98049343; PMID:9389475
A;Accession: D69361
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-212 <KLE>
A;Cross-references: GB:AE001042; GB:AE000782; NID:g2689365; PIDN:AAB90348.1; PID:g264
C;Superfamily: conserved hypothetical protein MJ1501

```
  Query Match             6.0%;  Score 142;  DB 2;  Length 212;
  Best Local Similarity   25.4%;  Pred. No. 0.0006;
  Matches   49;  Conservative   44;  Mismatches   84;  Indels   16;  Gaps    6;

Qy      37 GSGDFAKSLTIRLIRCGYHVVIGSRNPK----FASEFFPHVVDVT----HHEDALTKTNI 88
           |:|:  | | :||   |: :|:|||  :     |:|:     | :     :|||    :|
Db       8 GTGNLGKGLALRLATLGHEIVVGSRREEKAEAKAAEYRRIAGDASITGMKNEDAAEACDI 67

Qy      89 IFVAIHREH-YTSLWDLRHLLVGKI----LIDVSNNMRINQY--PESNAEYLASLFPDSL 141
             : |  ||   :  ||:::|  ||    |: ||   :   |    | || :| :
Db      68 AVLTIPWEHAIDTARDLKNILREKIVVSPLVPVSRGAKGFTYSSERSAAEIVAEVLESEK 127

Qy     142 IVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIELARQLNFI-PIDLGSLSSARE 200
           :|   : : |       :     | :| :: :::: |: |  ::: : |:| | ||::|
Db     128 VVSALHTIPAARFANLDEKFDWDVPVCGDDDESKKVVMSLISEIDGLRPLDAGPLSNSRL 187

Qy     201 IENLPLRLFTLWR 213
           :|:|   :  : |
Db     188 VESLTPLILNIMR 200


RESULT 6
T10120
F420-dependent NADP reductase (EC 1.6.8.-) - Methanobacterium thermoautotrophicum (strai
C;Species: Methanobacterium thermoautotrophicum
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 20-Jun-2000
C;Accession: T10120
R;Berk, H.; Thauer, R.K.
FEBS Lett. 438, 124-126, 1998
A;Title: F420H2:NADP oxidoreductase from Methanobacterium thermoautotrophicum: identific
A;Reference number: Z16959; MUID:99037734; PMID:9821972
A;Accession: T10120
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-224 <BER>
A;Cross-references: EMBL:Y17210
A;Experimental source: strain Marburg
C;Superfamily: conserved hypothetical protein MJ1501
C;Keywords: oxidoreductase

  Query Match             6.0%;  Score 142;  DB 2;  Length 224;
  Best Local Similarity   29.2%;  Pred. No. 0.00064;
  Matches   59;  Conservative   29;  Mismatches   84;  Indels   30;  Gaps    8;

Qy      37 GSGDFAKSLTIRLIRCGYHVVIGSRNPKFASEFFPHVVDVTHHEDALTK--TN------- 87
           |:||    | :||   |  |:||||: : |      |:::   :|   |  ||
Db       8 GTGDQGLGLALRLALAGEEVIIGSRDAEKAVSAAQKVLEIAERDDLKVKGATNAEAAEEA 67

Qy      88 ---IIFVAIHREHYTSLWDLRHLLVGKILIDV----------SNNMRINQYPESNAEYLA 134
              |: | :  :  | |  ::  : ||:|||           |    |: :  | ||  |
Db      68 EVAILTVPLQAQMAT-LGSVKEAIKGKVLIDATVPIDSCLGGSAVRYIDLWDGSAAERAA 126

Qy     135 SLFPD--SLIVKGFNVVSAWALQ--LGPKDASRQVYICSNNIQARQQVIELARQLNFI-P 189
             |  : :   || :|| ||    || |      | |::  |     ||| ::: :
Db     127 RFLEDQGTRVAAAFNNISASALLDITGPVDC--DCLIASDHRDALDLASELAEKIDGVRA 184

Qy     190 IDLGSLSSAREIENLPLRLFTL 211
           || | | :|| || :   |  |
Db     185 IDCGGLENARVIEKITPLLINL 206


RESULT 7
D64487
hypothetical protein MJ1501 - Methanococcus jannaschii
C;Species: Methanococcus jannaschii
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C;Accession: D64487
R;Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A;Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C
A;Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A;Reference number: A64300; MUID:96337999; PMID:8688087
A;Accession: D64487
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-223 <BUL>
A;Cross-references: GB:U67591; GB:L77117; NID:g2826422; PIDN:AAB99514.1; PID:g1500389
C;Genetics:
A;Map position: REV1473617-1472946
C;Superfamily: conserved hypothetical protein MJ1501

  Query Match             5.7%;  Score 133;  DB 2;  Length 223;
  Best Local Similarity   21.9%;  Pred. No. 0.0033;
  Matches   46;  Conservative   52;  Mismatches   78;  Indels   34;  Gaps    7;

Qy      37 GSGDFAKSLTIRLIRCGYHVVIGSRNPKFASEFFPHVVDV-----------THHEDALT 84
           |:||    | :|| :    ::|||| : | |      ::              ::||
Db       8 GTGDQGFGLALRLAK-NNKIIIGSRKKEKAEEAAKKAKEILKQRGIEADIIGLENKDAAK 66

Qy      85 KTNIIFVAIHREH-YTSLWDLRHLLVGKILIDV---------SNNMRINQYPE-SNAEYL 133
           : ::: :::  |:  :::  |:  | |||:: :             |:   |: | || :
Db      67 EGDVVILSLPYEYTLSTIKQLKEELKGKIVVSIGVPLATAIGDKPTRLLFPPDGSVAEMV 126

Qy     134 ASLFPDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIELARQLNFI-PIDL 192
             ::  :| :|  |  |    |:         : :| |: :|:: ||:|| |:: :  ||
Db     127 QNVLKESKVVSAFQNVCHAVLEDLDNPVDCDILVCGNDEEAKKVVIDLANQIDGVRAIDC 186

Qy     193 GSLSSAREIENLPLRLFTLWRGPVVVAISL 222
           |:|  :| || :          |::: :::
Db     187 GNLEKSRIIEAIT---------PLLIGLNI 207


RESULT 8
F86826
hypothetical protein yqfE [imported] - Lactococcus lactis subsp. lactis (strain IL140
C;Species: Lactococcus lactis subsp. lactis
C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C;Accession: F86826
R;Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Eh
Genome Res. 11, 731-753, 2001
A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis
A;Reference number: A86625; MUID:21235186; PMID:11337471
A;Accession: F86826
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-191 <STO>
A;Cross-references: GB:AE005176; PID:g12724621; PIDN:AAK05712.1; GSPDB:GN00146
A;Experimental source: strain IL1403
C;Genetics:
A;Gene: yqfE

  Query Match             5.1%;  Score 121;  DB 2;  Length 191;
  Best Local Similarity   22.2%;  Pred. No. 0.024;
  Matches   47;  Conservative   33;  Mismatches   74;  Indels   58;  Gaps    7;

Qy      32 TVGVIGSGDFAKSLTIRLIRCGYHVVIGSRNPKFASEFFPHVVDVTHHEDALTKTN---- 87
           |: : | |   |::                        |   |  | :     :||
Db       3 TISIFGKGKMGKAI-------------------GDNFSSSVNKVNYILSNSSKTELGEI 42

Qy      88 -------IIFVAIHREHYTSLWDLRHLLVGKILIDVSNNMRINQYPE-------SNAEYL 133
                   :   | :|: |        | |||:||::| :    :          | |  :
Db      43 VVLAVPYVAIAGIIQEYSTD-------LQGKIIIDITNPVDFTTFDSLLVPSDTSAAALI 95

Qy     134 ASLFPDSLIVKGFNVVSAWALQLGPKDASRQ----VYICSNNIQARQQVIELARQLNFIP 189
           |   |:|:||| ||  : ::  |  |  : :    | : |:: :|:: :|:
Db      96 AKQLPNSMIVKAFN--TTFSDTLATKKVANEHQTTVLLASDSQEAKETIIKALENSGLSL 153

Qy     190 IDLGSLSSAREIENLPLRLFTL-------WRG 214
           :| |||  |||:| :      ||       | |
Db     154 LDAGSLKRARELEAIGFLQITLAASEKISWDG 185


RESULT 9
A69131
```

conserved hypothetical protein MTH248 - Methanobacterium thermoautotrophicum (strain Del
C;Species: Methanobacterium thermoautotrophicum
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 29-Sep-1999
C;Accession: A69131
R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct
A;Reference number: A69000; MUID:98037514; PMID:9371463
A;Accession: A69131
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-232 <MTH>
A;Cross-references: GB:AE000811; GB:AE000666; NID:g2621287; PIDN:AAB84754.1; PID:g262129
A;Experimental source: strain Delta H
C;Genetics:
A;Gene: MTH248
A;Start codon: TTG
C;Superfamily: conserved hypothetical protein MJ1501

Query Match 5.1%; Score 120.5; DB 2; Length 232;
Best Local Similarity 23.8%; Pred. No. 0.034;
Matches 49; Conservative 42; Mismatches 90; Indels 25; Gaps 6;

```
Qy      30 KVTVGVIGSGDFAKSLTIRLIRCGYHVVIGSRNPKFASEFFPHVVDVTHHE--------- 80
           |: | : |:||    | :|    |  |:||||: : ||:    |:::   :
Db       8 KIAV-IGGTGDQGLGLALRFAVAGEEVIIGSRDAEKASKAASKVLEIAGRDDISVEGATN 66

Qy      81 -DALTKTNIIFVAIH-REHYTSLWDLRHLLVGKILIDVS---------NNMRINQYPESN 129
            ||    ::: : :  :    :|  :|  :  |:||| :          : :|     | :
Db      67 PDAAASADVVVLTVPLQAQMVTLASIRDQVRDKVLIDATVPIDSCIGGSAVRYIDLWEGS 126

Qy     130 AEYLASLF---PDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIELARQLN 186
           |   |: |     : :   || :|| ||    :       : |::  | :   ||| :::
Db     127 AAERAARFLREQGTRVAAAFNNISASALLEVSEPVDCDCLVASDHRDALEVAAELAEKID 186

Qy     187 FI-PIDLGSLSSAREIENLPLRLFTL 211
           :  |: | | :|| || :   |  |
Db     187 GVRAIECGGLENARIIEKITPLLINL 212
```

RESULT 10
T00121
hypothetical protein 8 - Leptospira interrogans
C;Species: Leptospira interrogans
C;Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 21-Jul-2000
C;Accession: T00121
R;Takahashi, Y.; Akase, K.; Hirano, H.; Fukunaga, M.
Gene 215, 37-45, 1998
A;Title: Physical and genetic maps of the Leptospira interrogans serovar icterohaemorrha

A;Reference number: Z14115; MUID:98332717; PMID:9666070
A;Accession: T00121
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-216 <TAK>
A;Cross-references: EMBL:AB010203; NID:g2780763; PIDN:BAA24368.1; PID:g2780765
A;Experimental source: strain Ictero No.1; substrain icterohaemorrhagiae

Query Match 5.1%; Score 119.5; DB 2; Length 216;
Best Local Similarity 23.7%; Pred. No. 0.037;
Matches 47; Conservative 37; Mismatches 91; Indels 23; Gaps 6;

```
Qy      33 VGVIGSGDFAKSLTIRLIRCGYHVVIGSRNPKFASEFFPHV---VDVTHHEDALTKTNII 89
           :|:::|||   ::|    :: |  | ||:|:     ::         :    :|     ||
Db       6 IGILGSGIVGQTLANGFLKYGAEVKIGTRDFGKLKDWLAKAGAGASIGSFSEAANFGEII 65

Qy      90 FVAIHREHYTSLWDLRHL--LVGKILIDVSNNMR----------INQYPESNAEYLASL 136
           :       : :  |  :  | || :|| :| :                | ||  | |
Db      66 VLCSKGSVASEVLTLSGIDSLNGKTIIDTTNPISEIPPQNGVLNFFTSYNESLMEKLQKQ 125

Qy     137 FPDSLIVKGFNVVSAWALQLGP--KDASRQVYICSNNIQARQQVIELARQLNFIPIDLGS 194
```

```
           | :  || |: | :  | : |  |     ::|| |:  :::|: |:   |  :   |:|
Db     126 APKANFVKCFSSVGS-GLMVNPQLKGEKPSMFICGNDDSSKKQIKEILDTFGWDTEDMGK 184

Qy     195 LSSAREIENLPLRLFTLW 212
           : :|| || | :   ||
Db     185 VEAARAIEPLCI----LW 198
```

RESULT 11
G82642
conserved hypothetical protein XF1737 [imported] - Xylella fastidiosa (strain 9a5c)
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C;Accession: G82642
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717; PMID:10910347
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: G82642
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-242 <SIM>
A;Cross-references: GB:AE003997; GB:AE003849; NID:g9106805; PIDN:AAF84546.1; GSPDB:GN
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.
, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv
M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF1737

Query Match 5.0%; Score 116.5; DB 2; Length 242;
Best Local Similarity 24.0%; Pred. No. 0.075;
Matches 46; Conservative 40; Mismatches 85; Indels 21; Gaps 7;

```
Qy      28 ARKVTVGVIGSGDFAKSLTIRLIRCGYHVVIGSRNP----KFASEFFPHVVDVTHHEDAL 83
           |  : :||||:|    ::    :: |: |:  ||||      | |  |    |     |
Db      30 AAPMRIGVIGAGSLGGTVGRLWVKAGHEVMFSSRNPDKLEAMARELEPR-ASVGQPLAAT 88

Qy      84 TKTNIIFVAIHREHYTSLW-DLRHLLVGKILIDVSNNMRINQ---YPESN----AEYLAS 135
              :: :|:  |    :  |||    |||::| :|    :    | |:     |: :
Db      89 EFGTVLLLAVPFEALPQVGRDLRSAYRGKIVLDSTNPWGASSADVYREARELGVAQTVVK 148

Qy     136 LFPDSLIVKGFNVVSAWALQLGPKDASRQ-----VYICSNNIQARQQVIELARQLNFIPI 190
           | : :|: |: | |  ::     |||:     : : |:: :| :    | |     |:
Db     149 YMPGARLVRAFSAVDATVVE---TSASRRGGRIGMPLASDDAEAMKVAEGLVRDAGCDPV 205

Qy     191 DLGSLSSAREIE 202
           :|:|::|   :
Db     206 IVGNLAAAASFQ 217
```

RESULT 12
AB3182
conserved hypothetical protein Atu5183 [imported] - Agrobacterium tumefaciens (strain
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002_
C;Accession: AB3182
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo
erage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McCl
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001

```
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; PMID:11743193
A;Accession: AB3182
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-198 <KUR>
A;Cross-references: GB:AE008687; PIDN:AAL45872.1; PID:g17743615; GSPDB:GN00188
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu5183
A;Genome: plasmid

  Query Match             4.9%;  Score 116;  DB 2;  Length 198;
  Best Local Similarity   22.7%;  Pred. No. 0.064;
  Matches   50;  Conservative   41;  Mismatches   81;  Indels   48;  Gaps   10;

Qy      31 VTVGVIGSGDFAKSLTIRLIRCGYHVVI-GSRNPKFASEFFPHV---VDVTHHEDALTKT 86
           :|||:||:|:    :    | : |   ||  || |:  :     :   :      :|  :
Db       1 MTVGIIGAGNIGGAFATALGKAGIEAVIANSRGPESLTALVSKLGSTIRAGSVPEAAAQA 60

Qy      87 NIIFVAIHREHYTSLWD-----LRHLLVG-KILIDVSNNMRINQY------PESNAEYLA 134
             |: ||:        |      |  |  | :|:|| :|::    |        :: :
Db      61 -IVLVAVP-------WSKIPGALAGLNFGDRIVIDANNSIEAPLYRPADLGGRTSTDIFT 112

Qy     135 SLFPDSLIVKGFNVVSAWALQLGPKDAS---------RQVYICSNNIQARQQVIELARQL 185
           :| | : :|| ||         | ||  |         | ::   ::::|: :|  :  ::
Db     113 ALVPGARVVKAFN-------HLTPKQLSGDPHSEGGRRVLFYSGDDMRAKAEVGAIIDRI 165

Qy     186 NFIPIDLGSLSSAREIENLPLRLFTLWRGPVVVAISLATF 225
             |  |||| |    ::   |        |  : |::|  |
Db     166 GFFGIDLGGLPVGSQLLQFP--------GGPLPALNLVKF 197


RESULT 13
JN0898
follitropin receptor precursor - crab-eating macaque
N;Alternate names: follicle-stimulating hormone receptor (FSHR)
C;Species: Macaca fascicularis (crab-eating macaque)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: JN0898; S36452
R;Gromoll, J.; Dankbar, B.; Sharma, R.S.; Nieschlag, E.
Biochem. Biophys. Res. Commun. 196, 1066-1072, 1993
A;Title: Molecular cloning of the testicular follicle stimulating hormone receptor of th
A;Reference number: JN0898; MUID:94071854; PMID:7504463
A;Accession: JN0898
A;Molecule type: mRNA
A;Residues: 1-695 <GRO>
A;Cross-references: EMBL:X74454; NID:g396801; PIDN:CAA52463.1; PID:g396802
A;Note: the authors translated the codon AGT for residue 488 as Arg
C;Function:
A;Description: receptor that mediates the biochemical effects of follitropin
C;Superfamily: glycoprotein hormone receptor; leucine-rich alpha-2-glycoprotein repeat h
C;Keywords: G protein-coupled receptor; glycoprotein; hormone receptor; phosphoprotein;
F;1-17/Domain: signal sequence #status predicted <SIG>
F;18-695/Product: follitropin receptor #status predicted <PFH>
F;71-95/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR2>
F;96-120/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR3>
F;121-145/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR4>
F;146-169/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR5>
F;172-193/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR6>
F;194-218/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR7>
F;367-387/Domain: transmembrane #status predicted <TM1>
F;399-421/Domain: transmembrane #status predicted <TM2>
F;444-465/Domain: transmembrane #status predicted <TM3>
F;486-508/Domain: transmembrane #status predicted <TM4>
F;529-550/Domain: transmembrane #status predicted <TM5>
F;574-597/Domain: transmembrane #status predicted <TM6>
F;609-630/Domain: transmembrane #status predicted <TM7>
F;191,199,293,318/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;555/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F;596/Binding site: phosphate (Ser) (covalent) (by protein kinase C) #status predicted

  Query Match             4.7%;  Score 111.5;  DB 1;  Length 695;
  Best Local Similarity   17.0%;  Pred. No. 0.71;
  Matches   84;  Conservative   75;  Mismatches  151;  Indels  183;  Gaps   16;

Qy      79 HEDALTKTNIIFVAIHREHYTSLWDLRHLLVG----------------KILIDVSNNMR 121
           ||  : | | : : |: | : :| :||:||:                  |:|:|: :|:
Db      98 HEIRIEKANNL-LYINPEAFQNLPNLRYLLISNTGIKHLPDVHKIHSFQKVLLDIQDNIN 156

Qy     122 INQYPESNAEYLASLFPDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIEL 181
           |:           ::  :| :   |  |  |      |:  :::: |:
Db     157 IH----------TIERNSFVGLSFESVILWL----NKNGIQEIHNCA------------ 189

Qy     182 ARQLNFIPIDLGSLSSAREIENLPLRLFTLWRGPVVVAISLATFFFLYSFVRDVIHPYAR 241
               |   :|  :||    :| ||  :|    |||:: ||      | |:  : :
Db     190 ---FNGTQLDELNLSDNNNLEELPNDVFHGASGPVILDISRTRIHSLPSYGLENLKKLRA 246

Qy     242 NQQSDFYKIPIEIVNKTLPIVAITLLSLVY------------------------------ 271
             :  |:|         :||:   || |
Db     247 RSTYNLKKLP-----SLEKLVALMEASLTYPSHCCAFANWRRQISELHPICNKSILRQEV 301

Qy     272 ------------LAGLLAAAYQLYYGTKYRRF-------------PPWLETWLQCRKQLG 306
                       ||    ::|   :  |  |               |  : :  |   ||
Db     302 DYMTQTRGQRSSLAEDNESSYSRGFDMTYAEFDYDLCNEVVDVTCSPKPDAFNPCEDILG 361

Qy     307 -----LLSFFFAMVHVA-------------YSLCLPMRRSERYLFLNMAYQQVHANIENS 348
               :| :| ::: :                | | :|     |:|  |:|:
Db     362 YNILRVLIWFISILAITGNIIVLVTLTTSQYKLTVP-----RFLMCNLAFAD-------- 408

Qy     349 WNEEEVWRIEMYISFGIMSLGLLSLLAVTSIPSVSNALNWRE---------FSFIQSTLG 399
                  :  ||  | : |:   |      : |::|:          |:   | |
Db     409 ------------LCIGIYLLLIASVDIHTKSQYHNYAIDWQTGAGCDAAGFFTVFASELS 456

Qy     400 ---YVALLISTFH-------------------VLIYGWKRAFEEEYYRFYTPPNFV-LAL 436
               |: :  :|                   |:: ||  ||    :  :   ::: :::
Db     457 VYTLTAITLERWHTITHAMQLDCKVHVRHAASVMVMGWIFAFAAALFPIFGISSYMKVSI 516

Qy     437 VLPSIVILDLLQL 449
           ||  :   | ||
Db     517 CLPMDIDSPLSQL 529


RESULT 14
S59681
probable membrane protein YPL012w - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein LPA5w; hypothetical protein YP8132.01
C;Species: Saccharomyces cerevisiae
C;Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 19-Apr-2002
C;Accession: S59681; S52519
R;Hall, J.; Ahmed, A.; Bussey, H.; Fortin, N.; Friesen, J.D.; Storms, R.K.; Vo, D.H.;
submitted to the EMBL Data Library, August 1995
A;Description: The sequence of Saccharomyces cerevisiae chromosome XVI left arm.
A;Reference number: S59677
A;Accession: S59681
A;Molecule type: DNA
A;Residues: 1-1228 <HAL>
A;Cross-references: EMBL:U33335; NID:g965076; PID:g965081; MIPS:YPL012w
R;Badcock, K.; Churcher, C.
submitted to the EMBL Data Library, February 1995
A;Reference number: S52519
A;Accession: S52519
A;Molecule type: DNA
A;Residues: 220-1228 <BAD>
A;Cross-references: EMBL:Z48483; NID:g683777; PID:g683778; MIPS:YPL012w
A;Experimental source: strain AB972
C;Genetics:
A;Gene: SGD:RRP12
A;Cross-references: SGD:S0005933
A;Map position: 16L
C;Keywords: transmembrane protein
F;125-141/Domain: transmembrane #status predicted <TM1>
F;478-494/Domain: transmembrane #status predicted <TM2>
F;720-736/Domain: transmembrane #status predicted <TM3>
```

```
  Query Match             4.7%;  Score 111.5;  DB 2;  Length 1228;
  Best Local Similarity   19.9%;  Pred. No. 1.5;
  Matches  111;  Conservative   71;  Mismatches  152;  Indels  225;  Gaps   28;

Qy      11 KSLSETCLPNGINGIKDARK------------------VTVGVIGSGDFAKSLTIRLI 50
Db     292 KTMAETTISSGLAENKYLRVLDTIFALKPSNVDTLLTKSWIAVVIKGMSTYATHQPLKAL 351

Qy      51 R---------CGYHVVIGSRNPKF---ASEFFPHVVDVTHHEDAL--------------- 83
Db     352 RKIPGVFHIMCTY---LASETPEVYQAASQCLISILSESVKDDLLLYTPSVDEKVFKNVD 408

Qy      84 ------TKTNIIFVAIHREHYTSLWDLRHLLVGKILIDVSNNMRINQYPESNAEYLASLF 137
Db     409 EIISQIAKTFIDFLSIRYSHCS-----REIL--KILVAAFNKFRY----RSNPHFLKSL- 456

Qy     138 PDSLIVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIELARQLNFI--------- 188
Db     457 ---------KIVDTWRVN------------------EEQFMDLRNEIELVIGASISAMG 488

Qy     189 --------PIDLGSLSSAREIENLPLRLFTLWRGPVV----VAISLATFFFLYSFVRDVI 236
Db     489 PEMILAEAPLNLDNPSSER-----PGR---AWLLPLIRDYTKNANLATF-------QNEL 533

Qy     237 HPYARNQQSDFYKIPIEIVN-----------KTLP--------------IVAITLLSL 269
Db     534 APYIKSFQSKFDKVPEESIQLRVFQTIVDQIWSTLPRFCELPMDLRESFTDEFASELSSL 593

Qy     270 VYL----------AGLLAAAYQLYYGTK--------YRRFP--------PWLETWLQCRK 303
Db     594 LYSEVELRTTICHALKVLAESNVSYAEESSSHNVLLLQRFPISEAQKNIEYLST-----K 648

Qy     304 QLGLLSFFFAMVHVAYSLCLPMRRS------ERYL------FLNMAYQQVHANIENSWNE 351
Db     649 STNLLAVLFNV----YTQTTPNARSYILETIDQYLKITSKEDLEKTFNNVCGLLKNSMNE 704

Qy     352 EEVWRIEMYISFGIMSLGLLSLL--AVTSIPSVSNALNWREFSFIQSTLGYVALLISTFH 409
Db     705 ESSGNVNKEKKKPQLTATLLDLIICMITYLPVSSYSALFSMFS----------LTVNSAD 754

Qy     410 VLIYGWKRAFEEEYYRFYT 428
Db     755 ALIQ--KRA-----YRIIT 766
```

RESULT 15
B84410
hypothetical protein Vng2607c [imported] - Halobacterium sp. NRC-1
C;Species: Halobacterium sp. NRC-1
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C;Accession: B84410
R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablo
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li
A;Title: Genome sequence of Halobacterium species NRC-1.
A;Reference number: A84160; MUID:20504483; PMID:11016950
A;Accession: B84410
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-222 <STO>
A;Cross-references: GB:AE004437; NID:g10581987; PIDN:AAG20646.1; GSPDB:GN00138
C;Genetics:
A;Gene: VNG2607C
C;Superfamily: conserved hypothetical protein MJ1501

```
  Query Match             4.7%;  Score 110.5;  DB 2;  Length 222;
  Best Local Similarity   24.3%;  Pred. No. 0.2;
  Matches   49;  Conservative   34;  Mismatches   94;  Indels   25;  Gaps    6;

Qy      37 GSGDFAKSLTIRLIRCGYH-VVIGSRNPKFASEFFPHVVDVT-----------HHEDAL 83
Db       8 GTGDIGAGLALRWATDSDHDIVIGSRDPEKARETAAAYEDTLADQGVDRKLTGFANEMAA 67

Qy      84 TKTNIIFVAIHREHYTSLW-------DLRHLLVGKILIDVSNNMRINQYPESNAE---YL 133
Db      68 DRADVVVVAVPAYHVTDVMGAVADRLDADTLVISPAVGIASGEHGLHYNPPSAGSVTALV 127

Qy     134 ASLFPDSL-IVKGFNVVSAWALQLGPKDASRQVYICSNNIQARQQVIELARQLNFI-PID 191
Db     128 ADAAPDGVDVVGAFHNLAADRLADLDTELDADTLVVGNDEGARTRVAELADDITGLRALD 187

Qy     192 LGSLSSAREIENLPLRLFTLWR 213
Db     188 AGPVENAAEVESLTPLLINLAR 209
```

Search completed: January 17, 2003, 19:00:16
Job time : 33 secs